Universitat Autònoma de Barcelona

Evolutionary ecology of range expansion in the colonizing plant Leontodon longirostris (Asteraceae)

Manuel de Pedro Rodríguez

Evolutionary ecology of range expansion in the colonizing plant *Leontodon longirostris* (Asteraceae)

Doctoral Thesis by:

Manuel de Pedro Rodríguez

Thesis Supervisors:

Dr. Maria Mayol Martínez Dr. Miquel Riba Rovira

February 2021

DOCTORADO EN ECOLOGÍA TERRESTRE

Departament de Biologia Animal, Biologia Vegetal i Ecologia,

Universitat Autònoma de Barcelona (UAB)

Centre de Recerca Ecològica i Aplicacions Forestals (CREAF)

Agradecimientos

Es bien conocido que la realización de una tesis doctoral no es el trabajo de una sola persona, sino que es el resultado de un esfuerzo colaborativo. Quiero aprovechar estas líneas para agradecer a todas las personas que la han hecho posible.

Primero de todo, gracias a mis directores, Maria Mayol y Miquel Riba, por darme la oportunidad de realizar esta tesis doctoral en su grupo. Gracias por toda la formación recibida a lo largo de estos años, por hacerme entender cómo se desarrolla el trabajo científico y que no todo es lo que parece a simple vista.

En segundo lugar, quiero agradecer a todas las personas sin cuya participación esta tesis no hubiera sido posible. Salvador y María Talavera compartieron con nosotros su inestimable saber acerca del género, y nos proporcionaron algunas de las muestras. ¡Perdonad si al final no seguimos vuestra propuesta taxonómica! Rocío Bautista, Manuel Gonzalo Claros y Pedro Seoane (Plataforma Andaluza de Bioinformática de la Universidad de Málaga) produjeron el borrador del genoma sobre el que se basaron los posteriores estudios genómicos. Carmen García-Barriga (INIA-CIFOR) llevó a cabo todas las extracciones de DNA, a pesar de la maldita tendinitis. Federica Cattonaro, Davide Scaglione y Simone Scalabrin (Istituto di Genomica Applicata, Udine) se encargaron de llevar a buen puerto la captura génica. La realización, mantenimiento y toma de medidas en los dos experimentos en condiciones controladas no hubiera sido posible sin los dolores de espalda de Ingrid Regalado, Jèssica Requena, Xavi Rico, José Miguel Sarradell y Guadalupe Márquez, que soportaron estoicamente el calor, las arañas y los picotazos de los mosquitos tigre. Luís López-Soria y Marc Estiarte compartieron desinteresadamente su tiempo y conocimientos (además de algún que otro esfuerzo en el invernadero) en largas discusiones sobre cómo llevar a cabo determinadas medidas. John R. Pannell, Delphine Grivet y Giovanni G. Vendramin siempre han estado ahí para compartir ideas e intereses, y para llevar a buen puerto el proyecto POREXPAN.

Muchas gracias también a Santi C. González Martínez (INRAE-BIOGECO) por acogerme en Burdeos durante mi breve estancia, por enseñarme los entresijos de los análisis en genética de poblaciones y por las charlas tan amenas con una sonrisa casi

perenne. También quiero agradecer la ayuda recibida por las compañeras con las que compartí despacho a un recién llegado sin entender nada de francés. Gracias, Katha y Marina.

Considero que el ambiente de trabajo es muy importante y en este aspecto he tenido suerte con los compañeros del CREAF. De forma especial quiero mencionar a los compañeros del despacho -150 con los que he compartido la mayor parte del tiempo y que han hecho que el trabajo sea motivador. Gracias, Judit, Javi G., Aina, Mari Ángeles, Víctor, Carlos, Marta y Pere. Quiero agradecer de forma especial a Judit por soportar las pesadas charlas y darme apoyo durante toda la etapa doctoral; Merci! Fuera del despacho también quiero agradecer a Kevin por sus prácticos consejos y a Laura por recalcar lo importante, que continuamente se nos olvida.

Creo que es justo mencionar los motivos que llevan a una persona a iniciar una tesis doctoral. En mi caso, se le tengo que agradecer a Mª Montserrat Martínez Ortega que, durante la realización de mi trabajo final de máster, me mostró con transparencia cómo era el mundo de la ciencia. Muchas gracias por los ánimos.

Aunque si soy sincero, el gusanillo de la ciencia me picó mucho antes. Recuerdo de forma especial la esfera Naukas, una plataforma de divulgación científica que despertó en mí la curiosidad por la ciencia (y el rechazo a las pseudociencias). Muchas gracias a todos los divulgadores que lo hicieron y lo siguen haciendo posible. ¡Larga vida a la «ciencia, escepticismo y humor»!

Fuera del ámbito de la ciencia y de la divulgación científica, no puedo olvidarme del apoyo recibido por Iván y Margaux durante el máster y etapas posteriores. Tampoco puedo dejar de mencionar a mis viejos amigos de Sayago que están ahí desde la infancia y con los que siempre puedo contar. Muchas gracias, Ana, Ara, Dani, Óscar y Raquel. Y aprovecho para darle la bienvenida a Ibai. Prometo que en el futuro nos veremos más a menudo.

Por supuesto, tengo que hacer una mención especial a mis padres, a los que he visto menos de lo que quería pero que han sido un pilar básico para haber llegado donde ellos no hubieran podido ni imaginar. Gracias por dejarme hacer lo que he querido. También quiero agradecer al numeroso entorno familiar. Gracias, Raquel, Alejandro y Rubén.

Por último, quiero agradecer de forma *muy* especial a Maite por todo su apoyo y aguante ante mis múltiples cambios de humor debido, como seguramente no puede ser de otra manera, a los vaivenes del doctorado. Pronto la distancia no nos separará. ¡Muchas gracias!

Esta tesis ha sido financiada por una ayuda para contratos predoctorales para la formación de doctores (BES-2015-074251) vinculada al proyecto de investigación POREXPAN (CGL2014-53120-P), ambas sufragadas por el Ministerio de Economía y Competitividad (Gobierno de España). Este trabajo también ha recibido soporte por parte de la Generalitat de Catalunya para fomentar las actividades de los grupos de investigación (2017SGR1006).

Table of Contents

General Introduction

A common observation about nature is that species are always restricted in space, and a central question of ecology and evolutionary biology is to understand the factors that govern the limits of species' geographical ranges (Holt & Barfield, 2011). This requires unraveling the interaction of different ecological and evolutionary processes acting at multiple levels of biological organization, from genes to communities (see Sexton et al., 2009; Holt & Barfield, 2011).

Historically, ecological factors have commonly been invoked to explain the limits of species' distributions. They have been summarized by Gaston (2003) quite a long time ago. Among them, different abiotic and biotic factors, in isolation or in combination, have been identified. Abiotic factors include physical barriers to dispersal that populations cannot overcome (e.g., mountain chains or rivers), and climatic conditions that surpass the physiological limits of species, preventing population persistence (e.g., high temperatures or low precipitation). Biotic factors limiting species range comprise interactions with other species, such as competitors, predators or parasites. In fact, the study and the relative importance given to all factors and processes governing species ranges has traditionally followed a somewhat hierarchical scheme. For instance, according to Gaston (2003), there are three broad questions that have to be addressed: what are the abiotic and biotic factors that mechanistically prevent further spread (e.g., physical barriers to dispersal), how these factors affect population dynamics and, finally, what is the genetic underpinning of species' responses to these factors. Probably, this hierarchical structure follows from the wealth of biogeographical and ecological observations and studies about the structure of range limits, i.e. on the form and position of range limits, and the shape of spatial distributions.

As it is nicely put by Kirkpatrick and Barton (1997), when physical barriers are the most important cause of range limits, range boundaries usually coincide with discontinuities in the habitat. In some other cases, the boundaries are more difficult to establish and appear more blurred and gradual. In such circumstances, the species range limits are essentially the expression of the species' ecological niche in space. Then, the question is, why the niche does not evolve in the range margins? From a conceptual point of view, the most common explanations to this question lie in evolutionary constraints, either genetic drift or the low rate of mutational input in small marginal populations or gene swamping (Bridle & Vines, 2007). However, this somehow static (equilibrium) perspective does not fully take into account that species ranges are highly

mobile, often shifting, expanding and contracting over time (Sexton et al., 2009), particularly during periods of large and rapid climate shifts. Because of the unprecedented rates of climate change, many organisms are expected to shift their ranges, and the evolutionary consequences of range shifts, and particularly the effect of range expansion, has recently become a major focus of both theoretical and empirical study.

The importance of range expansions

Range expansions are a ubiquitous aspect of species' natural history, and all species, at some point, have expanded to the site they currently occupy (Davis et al., 2005). It is well known that species have experienced major changes in their geographical distribution over the last million years. In Europe, for example, warming periods during the Pleistocene have allowed many species to expand their ranges northward from southern glacial refugia in the northern hemisphere (Hewitt, 1999).

Range expansions are currently occurring at an increasing rate in response to rapid climate change (Parmesan, 2006; Chen et al., 2011). Several species have shifted their ranges on average 17.6 km per decade towards higher latitudes and 12.2 m per decade to higher altitudes (Chen et al., 2011). Range expansions also play a determinant role in biological invasions, a significant component of global change, because the severity of the invasion is influenced by the spatial expansion after the introduction of non-native species into new habitats (e.g., Estoup et al., 2004).

Therefore, the success or failure to expand will determine where a species occurs and where it does not, which is directly implicated in the formation of species ranges (Bridle & Vines, 2007; Sexton et al., 2009). In some cases, such as in species expanding along environmental gradients, successful expansion may require the ability to adapt to the novel selection pressures encountered (e.g., Davis & Shaw, 2001; Thomas et al., 2001).

Understanding the implications of range expansions on species' ecology, genetics and biology is thus a relevant topic in evolutionary biology which allows better predictions of how species will respond to the ongoing global change (Mustin et al.,

2009), spread of pathogens (Brockmann & Helbing, 2013), agricultural pests (Lee et al., 2007) or invasive species (Alexander & Edwards, 2010).

In recent years, a growing number of theoretical studies are emerging investigating the characteristics and consequences of range expansions (Klopfstein et al., 2006; Excoffier et al., 2009; Burton et al., 2010; Phillips et al., 2010; Shine et al., 2011; Peischl et al., 2013; Gilbert et al., 2017; Peischl & Gilbert, 2020), and several experimental studies have tested such predictions in laboratory-based model systems, such as in bacteria (Hallatschek et al., 2008; Bosshard et al., 2017), insects (Ochocki & Miller, 2017; Szücs et al., 2017; Weiss-Lehman et al., 2017) and plants (Williams et al., 2016). However, few studies have examined these predictions in natural settings (in plants, González-Martínez et al., 2017; Laenen et al., 2018; Willi et al., 2018; Takou et al., 2019a), despite this is crucial because range expansions in nature may involve more complex interactions (see below) that may lead to different outcomes than those predicted by oversimplified models and tightly controlled laboratory settings.

Demography and genetics during range expansions

During range expansions, species are potentially exposed to new biotic and abiotic conditions against which they can respond adaptively (Davis & Shaw, 2001; Colautti & Lau, 2015). However, the spatial expansion itself might create particular demographic and genetic dynamics that may limit the adaptive capacity at the expanding edge. The most important theoretical demographic and genetic changes between the expanding populations (front) and non-expanding populations (core) are described below.

Demographic features during range expansions

Demographic features associated with range expansions are summarized in Fig. 0.1. Range expansions involve the establishment of groups of individuals into a new territory unoccupied by the species, usually by a small number of founders. In many cases, the successive founder events during expansion result in a gradient of decreasing population size along the colonization route (Phillips et al., 2010).

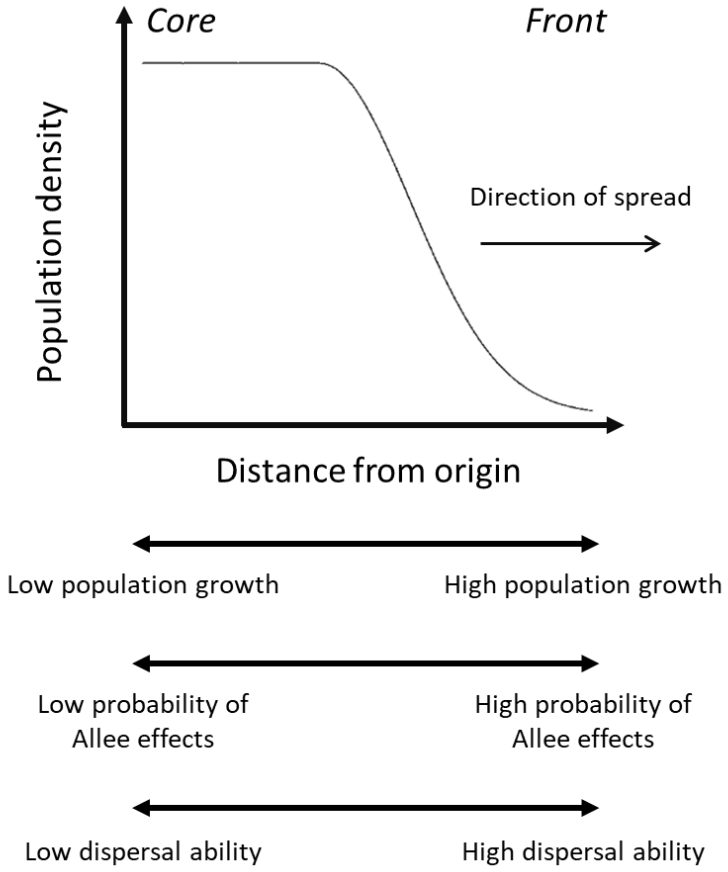


Figure 0.1. Demographic features at both ends of an expansion route. As the expansion proceeds, front populations experience lower conspecific density, higher population growth, higher probability of Allee effects and higher dispersal ability due to assortative mating (adapted from Phillips et al., 2010).

The low population size at the front-wave in the early stages of colonization generally implies low density. This gives rise to conditions where population growth is not constrained by conspecifics, resulting in higher growth rates in the front populations compared to those in the core (Phillips et al., 2010). In some occasions, the low conspecific density in populations at the expanding edge makes it difficult that sexual individuals find suitable mates, a phenomenon which is commonly referred to as Allee effect (Stephens et al., 1999). This effect is more severe in self-incompatible organisms and in plants that need animals for cross-pollination if they are not attracted by small populations (e.g., Groom, 1998; Elam et al., 2007).

Finally, individuals with greater dispersal ability are more likely to reach the expanding range edge and mate with each other (i.e., assortative mating by dispersal ability). This spatial sorting of individuals by dispersal ability along the expanding route

may result in highly dispersive individuals at the expanding front and low dispersal phenotypes remaining close to their natal locations (Shine et al., 2011).

Genetic consequences of range expansions

The demographic dynamics explained above may have important consequences on the patterns of genetic variation, in some cases potentially associated with adaptive responses brought about by the new biotic and abiotic conditions encountered in the new habitats. Repeated founder events, low densities, high growth and spatial sorting, all have a significant effect on the genetic patterns of expanding populations (see Fig. 0.2). As populations expand, the establishment of small founding populations cause the loss and random fixation of alleles through genetic drift, leading to an overall decrease of genetic diversity along the expansion route (Austerlitz et al., 1997).

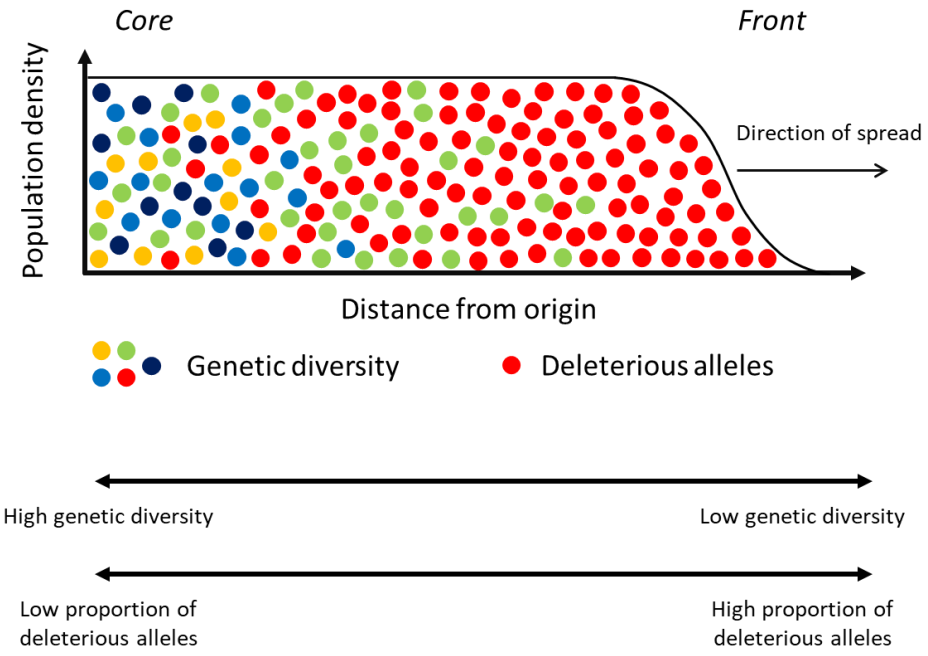


Figure 0.2. Genetic consequences of range expansions. The successive founder events lead to a reduced genetic diversity along the expansion route. The subsequent population growth after expansion may increase the frequency of some alleles that "surf" along the expansion. For instance, the red allele at the tip of the expanding range increases in frequency until becomes fixed at the expanding front (adapted from Excoffier & Ray, 2008).

On the other hand, the higher growth rate of populations at the colonizing front may increase the frequency of alleles that are not lost by genetic drift, which eventually can

be fixed through the whole population (i.e., "gene surfing"; Klopfstein et al., 2006). Gene surfing can affect either neutral, beneficial, or deleterious alleles. The surfing effect can lead to deleterious mutations reaching high densities at an expanding front, even when they have substantial negative effects on fitness (Travis et al., 2007). The mutation load associated with the accumulation of deleterious alleles via gene surfing during range expansion, the so-called "expansion load" (Peischl et al., 2013), may further reduce the ability to adapt to novel environments.

Trait changes associated with range expansions

Range-front populations may encounter new habitats along the expansion route, different from those at the core range, and then be subject to novel selection pressures. Therefore, traits favouring a species to colonize and expand in the range front may be different from those that allow a species to persist in the core region. In this section, we describe the types of traits that are expected to evolve during range expansions. We first consider functional traits, defined as those morphological or physiological traits which have an indirect impact on fitness (Violle, 2007). We then summarize life-history traits ("metatraits"), defined as those that are directly related to components of fitness, such as reproduction and survival (Braendle et al., 2011).

Functional traits

Individuals of populations at the expanding front, with initial low densities and subsequent periods of exponential population growth, are expected to be selected for increased size and growth. As an example, invasive plants tend to be bigger and grow faster than conspecifics from the native range (Siemann & Rogers, 2001; Keane & Crawley, 2002; Bossdorf et al., 2005). Similarly, the well-studied cane toads from recently expanded areas grow faster than toads from older populations (Phillips, 2009).

As mentioned before, assortment by dispersal ability drives the evolution of increased dispersal in front populations compared to those in the core range (Shine et al., 2011), a fact that has been well documented in cane toads (Phillips et al., 2006). In plants, an increase in dispersal ability after range expansion has been documented in natural settings in *Pinus contorta* (Cwynar & MacDonald, 1987), *Mycelis muralis* (Riba

et al., 2009) and in the invasive *Mikania micrantha* (Huang et al., 2015). Experimental studies also have demonstrated dispersal evolution associated with range expansion, in protists (Fronhofer & Altermatt, 2015), animals (Ochocki & Miller, 2017; Weiss-Lehman et al., 2017) and plants (Williams et al., 2016). The higher investment in growth and/or dispersal in expanding populations may decrease the investment in other traits, such as those related to defence against parasites or predators (Phillips et al., 2010). In general, however, colonizing individuals will mostly encounter generalist enemies and, therefore, will not need to invest in costly defences against specialized enemies (Müller-Schärer et al., 2004). In that sense, for example, plants from introduced populations tend to be better adapted to generalist herbivores than specialized ones (Joshi & Vrieling, 2005).

Physiological traits are also expected to change as a result of core-edge differences in selection pressures. Environmental conditions often change across space, such as day length, light quality or its seasonal variation, all of them often changing with latitude (Saikkonen et al., 2012). Changes in physiological traits have been found in invasive plants in Australia, which tend to have higher foliar nitrogen and phosphorus mass than native species, likely facilitating the invasion of nutrient-rich habitats (Leishman et al., 2007).

Life-history traits

The reduced density experienced by populations at the expanding front may allow to trade reproductive investment against investing in competitive ability (Burton et al., 2010). For instance, a greater reproductive output has been documented in invasive populations compared to their native counterparts (Siemann & Rogers, 2001; Keane & Crawley, 2002; Amundsen et al., 2012) and in experimental studies after range expansions (Szücs et al., 2017; Van Petegem et al., 2018).

On the other hand, in strictly alogamous species, individuals reproducing in populations at the expanding front may usually find few potential compatible mates and/or pollinators, a situation that may select for uniparental reproduction because self-fertilization provides reproductive assurance (Baker, 1955). This is often found during episodes of colonization (Pannell & Barrett, 1998; Rambuda & Johnson, 2004; Laenen et al., 2018). Despite selfing is often expected to evolve during expansions, self-

compatible organisms often express inbreeding depression, that is, fitness reduction due to the expression of deleterious recessive alleles, and a self-incompatible system may instead be favoured (see, for instance, Lande & Schemske, 1985; Barrett, 2014). In turn, high inbreeding depression and kin competition might select for increased dispersal (e.g., Hamilton & May, 1977; Roze & Rousset, 2005). Therefore, dispersal and mating system are closely linked components of life-history evolution.

One important component of life-history evolution, particularly in sessile organisms, is the timing of life cycle events (phenology). The timing of developmental events is especially important when adapting to new habitats, since it enables organisms to synchronize each life stage with the environmental conditions in which it can survive and reproduce. Therefore, traits related to phenology are expected to be subject to strong selection pressures during range expansion along environmental gradients (Griffith & Watson, 2006). Two important phenological traits in plants have been shown to be crucial for plant fitness: the timing of germination, which determines the conditions under which the seedling grow, and flowering time, that determines the conditions in which fertilization and seed maturation occur (Donohue et al., 2010).

The time to flowering has been widely documented to change in response to different environmental conditions during recent range expansions in native (Lustenhouwer et al., 2017) and in invasive plants (Weber & Schmid, 1998; Colautti et al., 2010). In general, studies show that many plant species have responded adjusting flowering time to optimize fitness according to the length of the growing season. (e.g., Weber & Schmid, 1998; Olsson & Agren, 2002; Parmesan & Yohe, 2003; Lustenhouwer et al., 2017). On the other hand, works studying the evolution of germination time have demonstrated that selection may favor early germination to take benefit of a longer growing season, possibly yielding reproductive advantages (e.g., Donohue, 2002; Shimono & Kudo, 2003; Castro, 2006). Early emergence may also provide a competitive advantage for limited resources against conspecific competitors that emerge later (e.g., Dyer et al., 2000). However, early-emerged seedlings may face a higher mortality risk due to seasonal hazards, such as pathogens, predation or desiccation (e.g., Marks & Prince, 1981).

Selection on phenological traits is expected to be particularly strong in short-lived monocarpic and annual plants, since their fitness is determined by a single reproductive event. Annual plants tend to delay seed germination in highly unpredictable

environments, e.g. arid habitats, until environmental conditions become favourable, and they can germinate and flower early to complete their cycle rapidly (Aronson et al., 1992; Volis et al., 2002; Manzano-Piedras et al., 2014; Vidigal et al., 2016). In contrast, annuals inhabiting more humid and temperate habitats may germinate in broader conditions and flower later since the environment is more predictable (Manzano-Piedras et al., 2014; Vidigal et al., 2016). Studies assessing the joint evolution of both germination and flowering time are rather limited (but see, for instance, Griffith & Watson, 2005; Prendeville et al., 2013; Galloway et al., 2018). Some exceptions include the model plant *Arabidopsis thaliana*, for which recent studies show that the timing of seed germination can have a large effect on flowering time (reviewed in Donohue et al., 2010). For instance, a study using populations from the Iberian Peninsula showed that south-western plants had higher seed dormancy and flowered early, suggesting adaptation to shorter growing seasons imposed by high temperatures (Vidigal et al., 2016).

Eco-evolutionary complexity during range expansions

Much of the theoretical predictions of range expansions outlined before, concerning the comparison between core and front populations, are often based on rather simplistic assumptions (homogeneous habitats, simple modes of trait inheritance, etc.). However, realistic scenarios involve greater complexity that can modify the expectations. This complexity arise from eco-evolutionary feedback mechanisms in which evolution can alter species demographical responses, and ecological processes may, in turn, modify evolutionary outcomes. The importance of eco-evolutionary feedbacks during expansions has been highlighted in several reviews (Phillips et al., 2010; Chuang & Peterson, 2016; Nadeau et al., 2019; Miller et al., 2020), and may probably account for current disagreement between theoretical and empirical studies.

For example, experimental studies showed that certainly bacterial populations evolved reduced fitness during expansion due to accumulation of expansion load, though the reduction in fitness was quite variable among experiments (Bosshard et al., 2017). In any case, empirical studies are rather limited, making it unclear whether such effects are common in nature. In humans, examples are mostly restricted to genomic comparisons between the Africa core region and the expanded populations in Asia and

the Americas, the latter showing a higher number of deleterious alleles, probably due to the successive bottlenecks during expansions (Henn et al., 2016). However, there is not a clear signal of their effects on fitness. In plants, for instance, front populations of *A. lyrata* subsp. *lyrata* showed both a higher proportion of deleterious alleles and a fitness reduction along the colonization route (Willi et al., 2018). However, the sister species *A. lyrata* subsp. *petraea* showed neither an accumulation of deleterious alleles after range expansion nor a negative effect on growth (Takou et al., 2019a). In this case, the discrepancies could be ascribed, at least in part, to different mating systems between species. While range-front populations of the former evolved towards selfing (Willi et al., 2018), populations of the latter maintained a self-incompatible system along the expansion route (Takou et al., 2019a), favouring gene flow and thus reducing the severity of founder events and expansion load.

The interplay between demography, breeding system and dispersal (gene flow) is particularly important for colonizing species. Since low densities are typically encountered at the edge of a population's range, fitness and expansion rates can decrease due to the difficulties in finding mates (i.e., Allee effects, Stephens et al., 1999). Several theoretical and observational studies suggest that evolution can mitigate an Allee effect. In the case of plants, restricted mate availability may select for selfing (e.g., Baker, 1955; Lande & Schemske, 1985). However, the evolution of selfing might be prevented by inbreeding depression (Lande & Schemske, 1985). Surfing of deleterious alleles during expansion may contribute to increase the negative effects of inbreeding depression in range-front populations, further preventing the evolution of selfing. On the other hand, theory predicts that dispersal is favored by high inbreeding (Perrin & Goudet, 2001; Roze & Rousset, 2005), thus alleviating the negative effects of Allee effects. Dispersal and, in general, gene flow, can sometimes mitigate the negative effects of deleterious mutations (Alleaume-Benharira et al., 2006; Sexton et al., 2011; Peischl & Gilbert, 2020). Similarly, theoretical models suggest that the evolution of dispersal, depending on the strength of resource competition, can speed up the spread rate even despite simultaneously exacerbating the Allee effect (Shaw & Kokko, 2015).

Besides the effects of drift, surfing and gene flow on the adaptive potential, traits do not occur in isolation. Instead, they are part of a complex and integrated phenotype that make up the whole organism, so that the change on one trait can modify another trait. Consequently, selection during range expansions will act not only on individual

traits, but also indirectly on multiple correlated traits (Lande & Arnold, 1983). As a consequence, evolution can be constrained by correlation among traits that are subject to antagonist selection pressures, even if sufficient genetic variation exists on these traits for natural selection to act (Etterson & Shaw, 2001). Some recent studies indicate that the sign, the magnitude and the type of correlation between the "metatraits" of survival, fertility and dispersal, generate quite different evolutionary results and rates of expansion (Ochocki et al., 2020). Currently, however, the role of genetic, maternal, and environmental correlations among traits during range expansion remain largely unexplored.

Combining phenotypic and population genetic approaches

Since natural selection drives changes both in phenotypic traits and the underlying allelic frequencies that partly determine them, the detection of selective pressures and their result, adaptation, can be evaluated both at the phenotypic and genetic level. Phenotypes are important since they are the target of natural selection, regardless of their genetic basis. Moreover, phenotypes integrate complex traits, such as phenology or growth, that might be the result of many genes or gene regulatory networks that interact with each other and with the environment. The action of selection on some traits is currently hard to detect using only genotypic approaches, since population genomic analysis (see below) can fail to detect adaptation in complex traits that are affected by subtle changes in allele frequency at many loci (polygenic selection), especially when genomic resources are limited (Wellenreuther & Hansson, 2016).

The most immediate problem in the practical study of phenotypic traits is to discern between the environmental and genetic contribution on the phenotype. Common garden experiments are a powerful way to test the genetic basis of trait variation by growing individuals from different populations in a common environment. Barring maternal effects (Mousseau & Fox, 1998; Wolf & Wade, 2009), in this experimental settings, which minimize the contribution of environmental variation, trait differences are expected to be due to the genetic differentiation. They are particularly useful when dealing with a large number of populations distributed along geographic and/or environmental clines (e.g., Mayr, 1956; Prendeville et al., 2013).

Population genomic approaches, on the other hand, can complement the evidence for selection and adaptation by searching signatures left by natural selection on the genome (Nielsen, 2005; Weigand & Leese, 2018). The neutral theory of molecular evolution posits that most genetic polymorphisms are either neutral or slightly deleterious, and that changes are mainly due to neutral forces (Kimura, 1968), whereas natural selection acts on specific loci. Identifying regions that seem exceptionally unusual compared with the rest of the genome is a successful approach for detecting genes or regions under selection (e.g., Fan et al., 2016; Mattila et al., 2016; Harpak et al., 2020). Genomic data can also be used to ascertain the dominant mode of selection which may vary between the core and the range fronts of the species' distribution.

A complete understanding of the process of natural selection also requires finding the environmental factor that acted as selective pressure giving rise to among-individual fitness differences and, eventually, leading to adaptation. Correlations between phenotypic or genotypic variability with the observed field environmental variability are indicative of the existence of adaptations associated with environmental clines (Hancock et al., 2011). However, some correlations may be the result of demographic process derived from range expansions and not by selection (Excoffier et al., 2009). In order to account for demographic history, a summary of among-individual relatedness (i.e., population structure) can be incorporated as a covariate. Even so, genotype-environment correlations are prone to false positives (De Mita et al., 2013). This is of particular concern in recent years where DNA genotyping has been more accessible to researchers due to the advent of next-generation sequencing technology. It thus appears that understanding selection will not be limited by the amount of genomic data, but by conceptual, analytical and phenotyping limitations. Combining several methods to detect signatures of selection at the molecular level greatly improves the quality of the inference.

Colonizing plants as models for studying range expansions

Colonizing plants are expected to experience frequent demographic variability, range contractions and expansions, often across environmental gradients requiring adaptive evolution (Colautti & Barrett, 2013). These characteristics make colonizing plants ideal for investigating ecological and evolutionary responses occurring during range

expansions. Furthermore, the short life cycles characterizing many colonizing plants make them amenable for experimentation.

The use of plant colonizers as models for the study of natural range expansions is not only of theoretical interest, but also of practical and applied interest, since many of them are invaders. Despite biological invasions are an important component of global change that threat biodiversity, ecosystem integrity, agriculture and human health (Mack et al., 2000), the invasive process is not well understood. Only a subset of introduced species prospers in establishing and only a few become widespread with a high impact on ecosystems (Williamson & Fitter, 1996). A determinant factor of the magnitude of damage caused by invasive species is the speed and extent to which these species expand (Epanchin-Niell & Hastings, 2010). Therefore, understanding the factors that favour range expansions in native colonizing species can be very useful to understand the invasion processes, offering practical guidelines to predict which introduced species will succeed in transitioning from establishment to spatial expansion and, in so doing, would become a potential invader.

Study species and area

The few empirical plant studies exploring the patterns and evolutionary consequences of range expansions are mostly restricted to the Brasicaceae family, either model plants or its relatives, such as *Arabidopsis thaliana* (Williams et al., 2016), *Arabidopsis lyrata* (Willi et al., 2018; Takou et al., 2019a), *Arabis alpina* (Laenen et al., 2018), *Eutrema salsugineum* (Wang et al., 2018) or *Capsella bursa-pastoris* (Kryvokhyzha et al., 2019). These studies give a valuable information but, as in many questions in ecology and evolution, in order to gain some generality it is necessary to explore other species with different demographic histories and life-history traits. At the same time, most of these studies use either a phenotypic or genotypic approximation. Here, we have selected the non-model plant species *Leontodon longirostris* (Asteraceae) as an experimental model to explore the evolutionary dynamics of range expansions, both at the phenotypic and genotypic levels.

Leontodon longirostris as a study species

Leontodon longirostris (Finch & PD Sell) Talavera (≡ *Leontodon saxatilis* subsp. *rothii* Maire ≡ *Thrincia hispida* Roth) is a widespread annual colonizing plant that belongs to the Asteraceae (Compositae) family, one of the largest family of plants (~10% of the world flora). The Asteraceae family harbours many colonizers, as evidenced by the lists of introduced species where the Asteraceae are widely represented, usually being the second most represented family after Poaceae (Pyšek, 1997).

Leontodon longirostris occurs in southern Europe, north-western Africa and the Macaronesian region (the Azores Islands, the Madeira Islands, the Salvage Islands, the Canary Islands and the Cape Verde Islands) (Ruiz de Clavijo, 2001) (Fig. 0.3A). It grows on cultivated and abandoned fields, roadsides, therophitic grasslands, and other disturbed spaces on a variety of soils (Talavera et al., 2015) (Fig. 0.3B). It can be found on a wide range of elevations, from sea level up to more than 1,500 m a.s.l. It is also widely naturalized in other regions with Mediterranean climate such as Chile, parts of the United States and southern Australia, where it is considered to be invasive (CGP; Groves et al., 2003).

Leontodon longirostris forms a basal rosette of leaves (Fig. 0.3C) with several unbranched scapes (leafless inflorescence stem), which end in a capitulum containing ligulate yellow flowers (Fig. 0.3D). In its natural range, flowering extends from March to August. It is a self-incompatible outcrossing species since self-pollinated flowers produce few, if any, achenes (dry one-seeded fruits) (Ruiz de Clavijo, 2001; García, 2004). It is pollinated by generalist insects, such as *Psilothrix nobilis* (Coleoptera), *Lachnaea* spp. (Coleoptera) or *Panurgus* spp. (Hymenoptera) (Ruiz de Clavijo, 2001). The number of capitula per plant ranges from 2 to 30, each one with 20 to 200 flowers (personal observation). It produces two types of fruits: central and peripheral achenes. Central achenes are produced in large numbers, are lighter and possess a well-developed pappus (with 10-14 bristles), while peripheral achenes are heavier and exhibit no pappus. These differences result in contrasting patterns of dispersal and germination (Ruiz de Clavijo, 2001). Central achenes show high spatial dispersal, germinate more rapidly and in higher proportion under broader conditions. By contrast, peripheral achenes disperse close to the mother plant and most of them remain in a dormant state forming a temporary seedbank until the germination conditions are favourable. The

differences between these fruit types are usually considered as an evolved bet-hedging strategy to cope with unpredictable environments (see Venable & Lawlor, 1980).

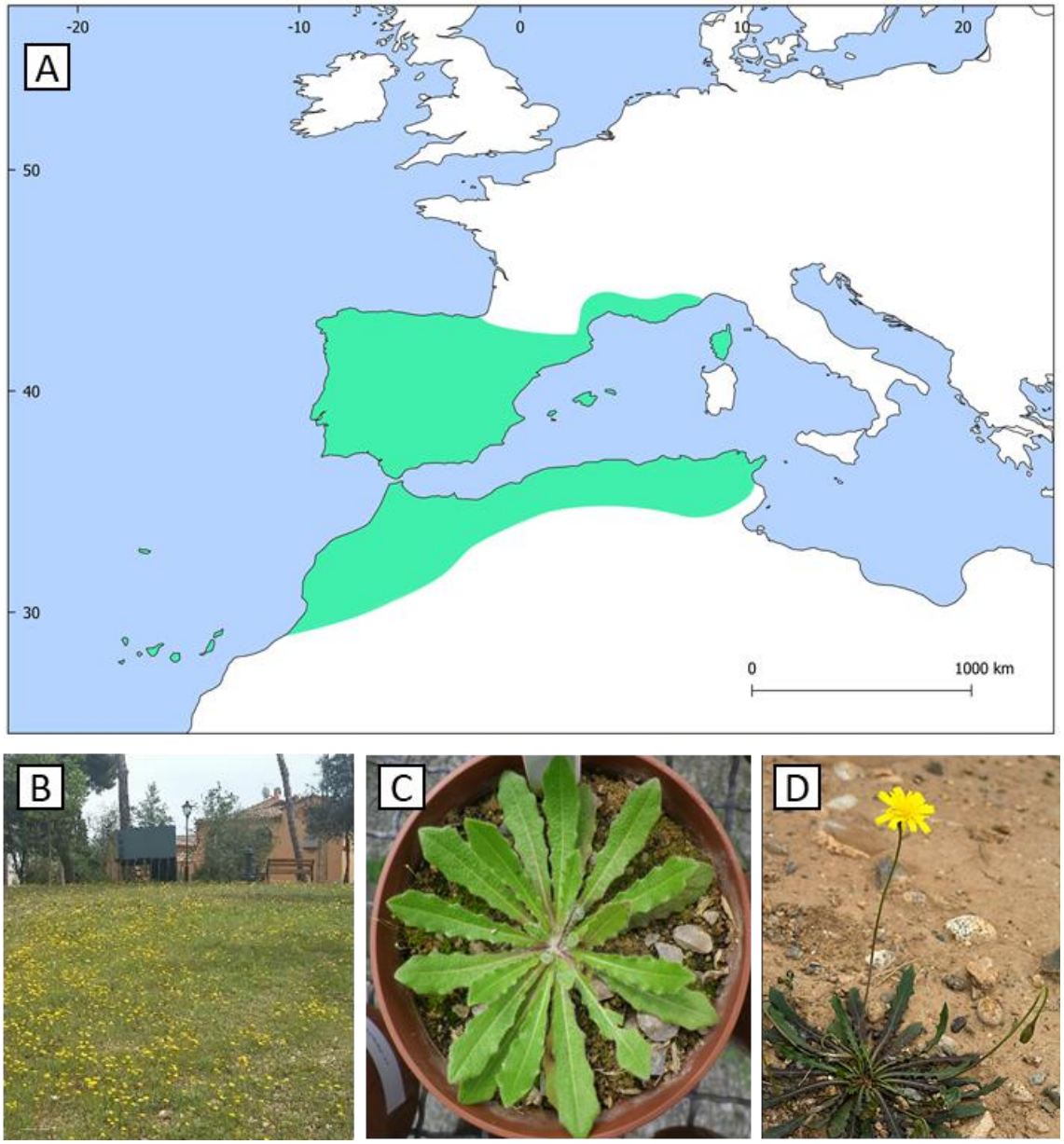


Figure 0.3. A) Map showing the approximate native distribution of *Leontodon longirostris* (green area, redrawn from Meusel & Jäger, 1992, incorporating information from Talavera et al., 2015; SIFlore dataset, Just et al., 2015); **B**) Habitat; **C**) Rosette; **D**) Individual with one flowering axis.

The ability of *L. longirostris* to grow on disturbed habitats, together with the short-lived cycle and high dispersal ability, suggests that it is a good colonizer plant that has probably experienced important and recent demographic changes and range shifts.

Its short-lived life cycle makes it amenable to study phenotypic traits using a common garden approach. Furthermore, it has one of the smallest genomes within the Asteraceae (Vallès et al., 2013), and a high-quality reference transcriptome assembly is already available for the sister species *L. saxatilis* Lam. (Hodgins et al., 2014), which greatly facilitates to perform population genomic studies.

Study area: the Iberian Peninsula

Leontodon longirostris is widely distributed within the Iberian Peninsula. It is mainly found in the area dominated by Mediterranean climate, which is characterized by mild rainy winters and hot dry summers with two peaks of maximum precipitation in spring and autumn (de Castro et al., 2005). However, the Mediterranean climate in the Iberian Peninsula is not a homogenous, and temperature (Fig. 0.4) and precipitation (Fig. 0.5) vary depending on the latitude and position versus the sea. Lower latitudes are warmer and drier than higher latitudes. While mean annual temperature often exceeds 18 ºC in the southern areas, lower temperatures are found in the north, with mean annual temperature below 15 ºC near the coast and below 10 ºC in inland areas. Similarly, annual precipitation can be less than 400 mm in the south, while it can often exceed 800 mm in northern areas (de Castro et al., 2005). In addition, the Iberian Peninsula is characterized by high interannual precipitation variability, particularly in southern latitudes (de Castro et al., 2005).

Hence, the Iberian Peninsula is a suitable setting to carry out evolutionary studies of range expansions across environmental gradients, usually arising across a complex matrix of spatially heterogeneous landscapes where ecological conditions often vary over short distances (Thompson, 2005).

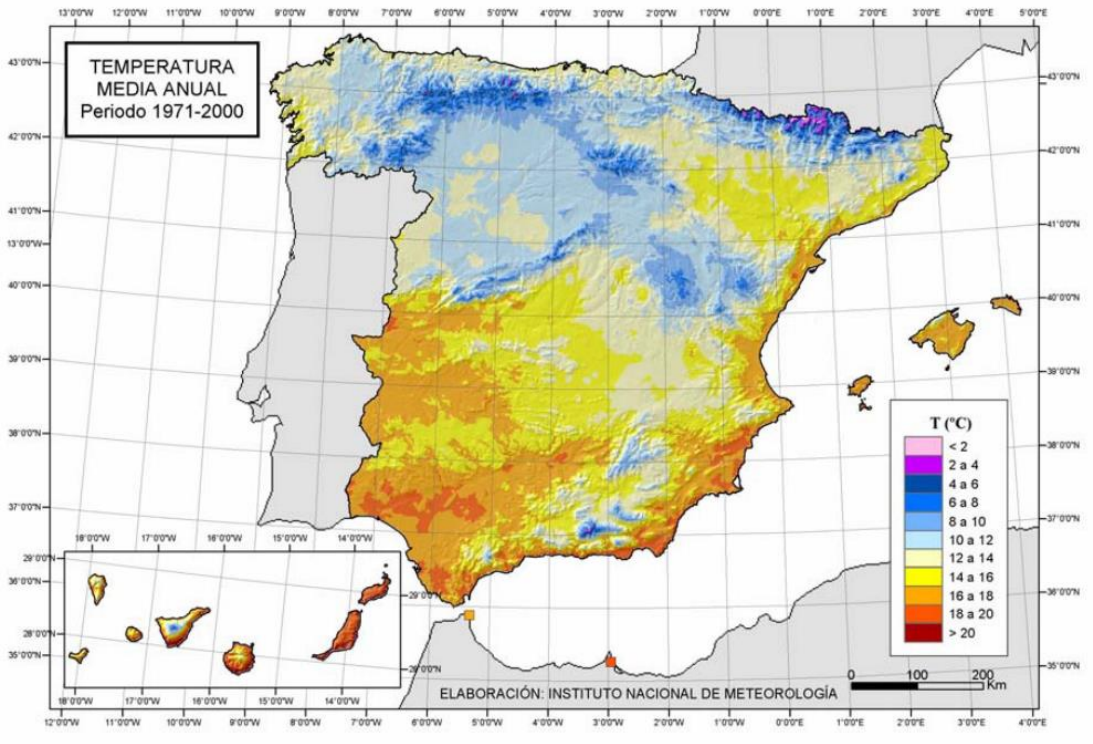


Figure 0.4. Average annual temperature (Cº) in Spain estimated from 1971 to 2000 (de Castro et al., 2005).

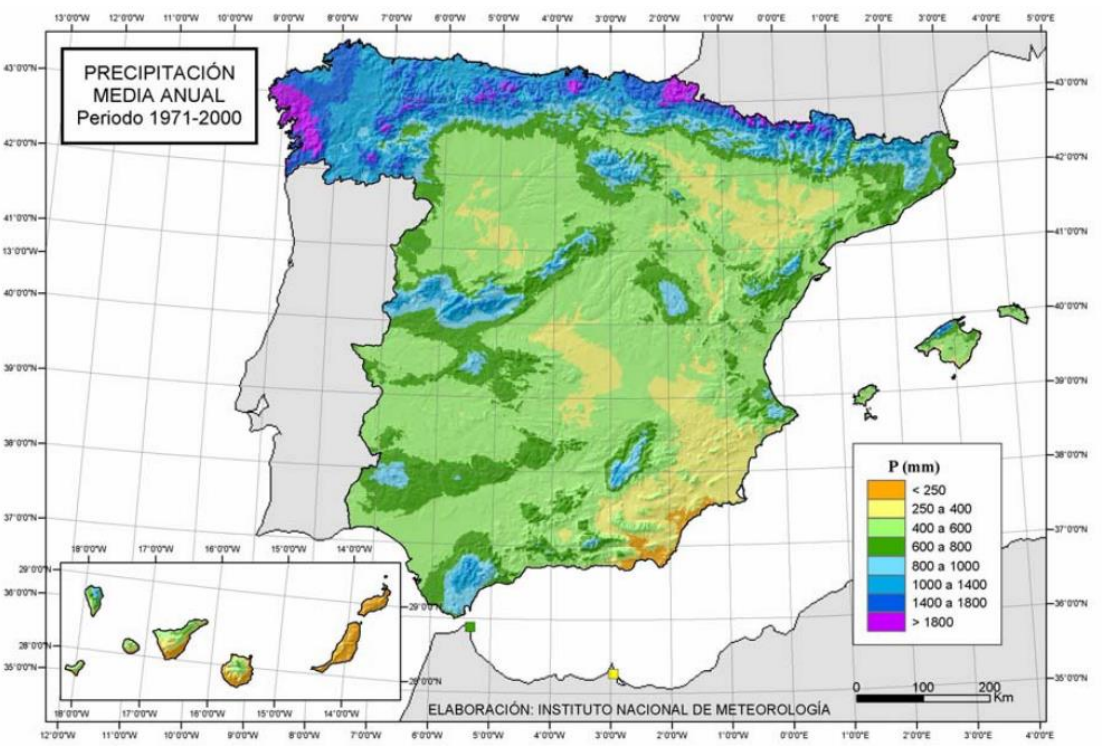


Figure 0.5. Average annual precipitation (mm) in Spain estimated from 1971 to 2000 (de Castro et al., 2005).

Objectives and structure of the thesis

The global aim of this thesis is to provide new experimental evidence and novel understanding on the role of evolutionary mechanisms during plant range expansions.

Using the annual colonizing plant *Leontodon longirostris* across most of its native range in the Iberian Peninsula, we first investigated demographic history of the species in the Iberian Peninsula (**Chapter 1**), which was dominated by a south-north range expansion. We then focused in the following specific objectives that are relevant to the evolutionary dynamics of expanded populations:

1) To evaluate the genetic consequences of the range expansion in terms of genetic diversity and expansion load (**Chapter 1**).
2) To ascertain whether the loss of genetic diversity and increased expansion load translate into lower adaptive capacity in range-front populations (**Chapters 2** and **3**).
3) To examine phenotypic differentiation among populations and (co)variation among life-history traits suggesting local adaptation, and to investigate the potential environmental factors that drive divergent selection and adaptation (**Chapter 2**).
4) To investigate the signatures of selection at molecular level, assessing whether the mode of selection differs between populations located at the core region and those at the front of the expansion wave (**Chapter 3**).
5) To identify candidate genes putatively under selection responsible for the observed trait variability and adaptation (**Chapter 3**).

The thesis is structured in three chapters that are summarized as follows:

Chapter 1. Demography, genetic diversity and expansion load.

In this chapter, we inferred the demographic history of *Leontodon longirostris* across most of its natural range in the Iberian Peninsula, and we examined the genetic

consequences of this expansion. Inferred demographic history supported a range expansion from southern Iberia around 40,000 years ago, reaching northern Iberia around 25,000 years ago. The expansion was accompanied by a loss of genetic diversity and a significant increase in the proportion of putatively deleterious mutations. However, levels of expansion load were smaller than those found in other plant species, which can be explained, at least partially, by its high dispersal ability, the self-incompatible mating system, and the fact that the expansion occurred along a strong environmental cline.

Chapter 2. **Environmental patterns of adaptation: phenology, size, and biomass allocation.**

Using common garden experiments, we assessed whether the reduction in genetic diversity and increased expansion load have impacted the adaptive ability of populations, looking for an integrated response of selection on several phenotypic traits. We investigated patterns of variation in phenological traits (the timing of germination and flowering) and determined its influence on other life-history traits linked to fitness (size at reproduction, biomass and reproductive allocation). We also analyzed these traits in the close relative *L. saxatilis*, which is a perennial lineage that can be found in more stable habitats. We found significant between and within species variability for most of the traits, with shorter lifespan linked to more unpredictable and stressful environments. These results strongly supported the presence of an adaptive cline for major life-history stage transitions, supporting that expansion load have not precluded local adaptation to the climatic variability encountered along the expansion route.

Chapter 3. Unraveling the genetic basis for adaptation during range expansion.

Finally, we integrated distinct genomic approaches with phenotypic information obtained from common garden experiments to better understand the potential adaptive processes that occurred during the northward expansion of *L. longirostris* in the Iberian Peninsula. As for the phenotypic traits, we found ample evidence that adaptation to new environments has occurred during the northward expansion, suggesting that (i) the loss of genetic diversity probably affected non-relevant (neutral) loci and (ii) expansion load

could be the result of hitchhiking of deleterious alleles with nearby positively selected variants. We also found ample evidence that adaptation during the expansion mainly proceeded by selection acting on genetic variation already present in the populations (i.e., standing genetic variation). Finally, we identified several candidate genes, mainly acting in signaling pathways, potentially mediating the life-history changes detected at the phenotypic level.

Methodology

The main methodological steps are summarized in Fig. 0.6. To achieve the previously stated objectives, several populations of *Leontodon longirostris* and the close relative *Leontodon saxatilis* were first sampled across the Iberian Peninsula (Fig. 0.7). Forty two populations of *L. longirostris* and six populations of *L. saxatilis* were grown in two common garden experiments for the analysis of phenotypic variation (**Chapter 2**). A subset of 21 populations of *L. longirostris* and 2 populations of *L. saxatilis* were used for genomic analyses (**Chapters 1** and **3**).

The analysis of molecular variation was based in the following steps. In collaboration with the Plataforma Andaluza de Bioinformática of the University of Málaga, we first generated a genome draft assembly and annotation for *Leontodon longirostris*, covering 418 Mbp (~53% of the genome). Then, we selected relevant genomic regions (potentially involved in local adaptation in other related Asteraceae) to design target capture probes (~1.5 Mbp) that were genotyped in 238 individuals of *L. longirostris* (21 populations) and 20 individuals of *L. saxatilis* (2 populations). After filtering, a total of 168,733 single nucleotide polymorphisms were retained.

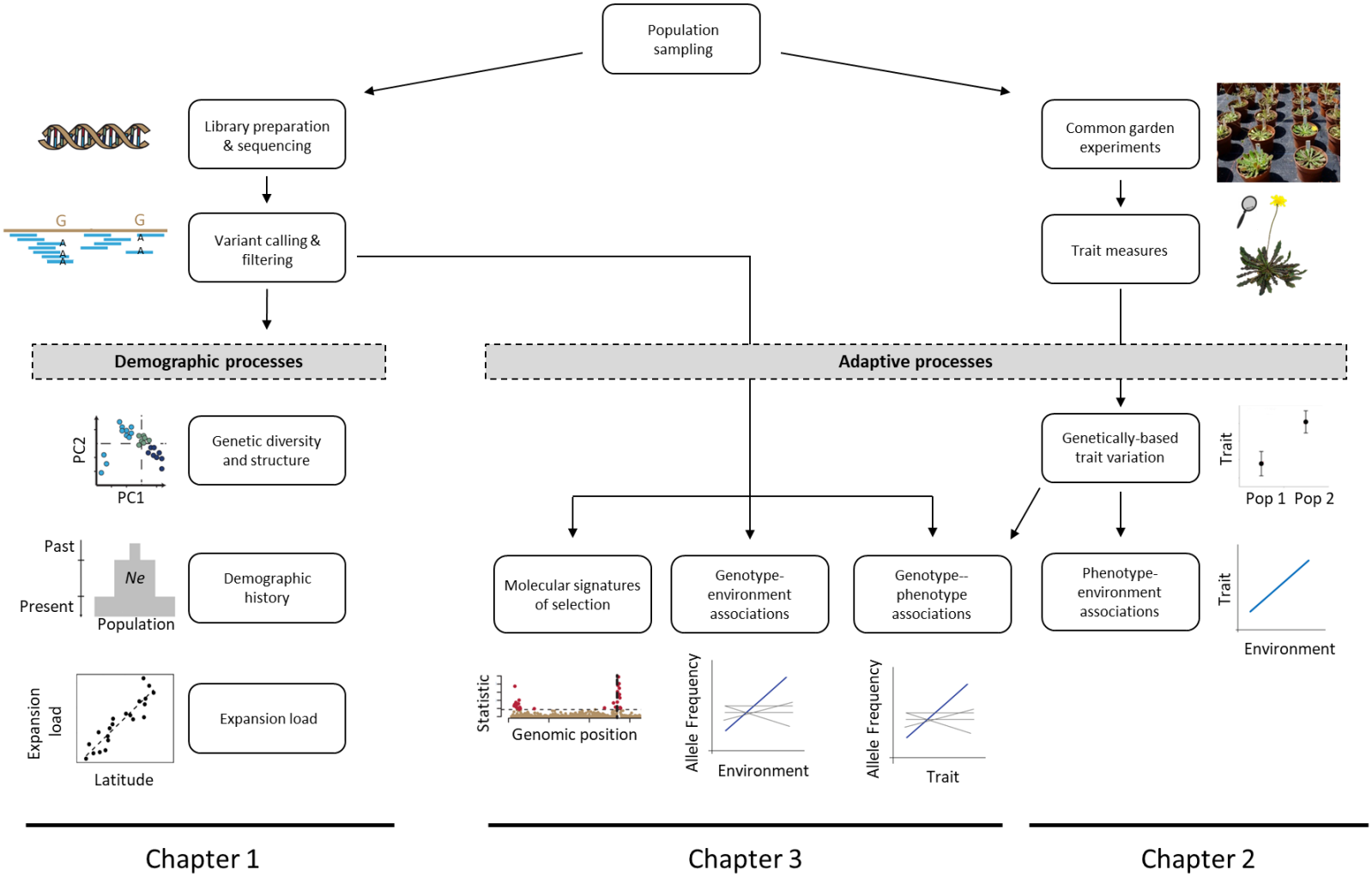


Figure 0.6. Workflow diagram of the thesis.

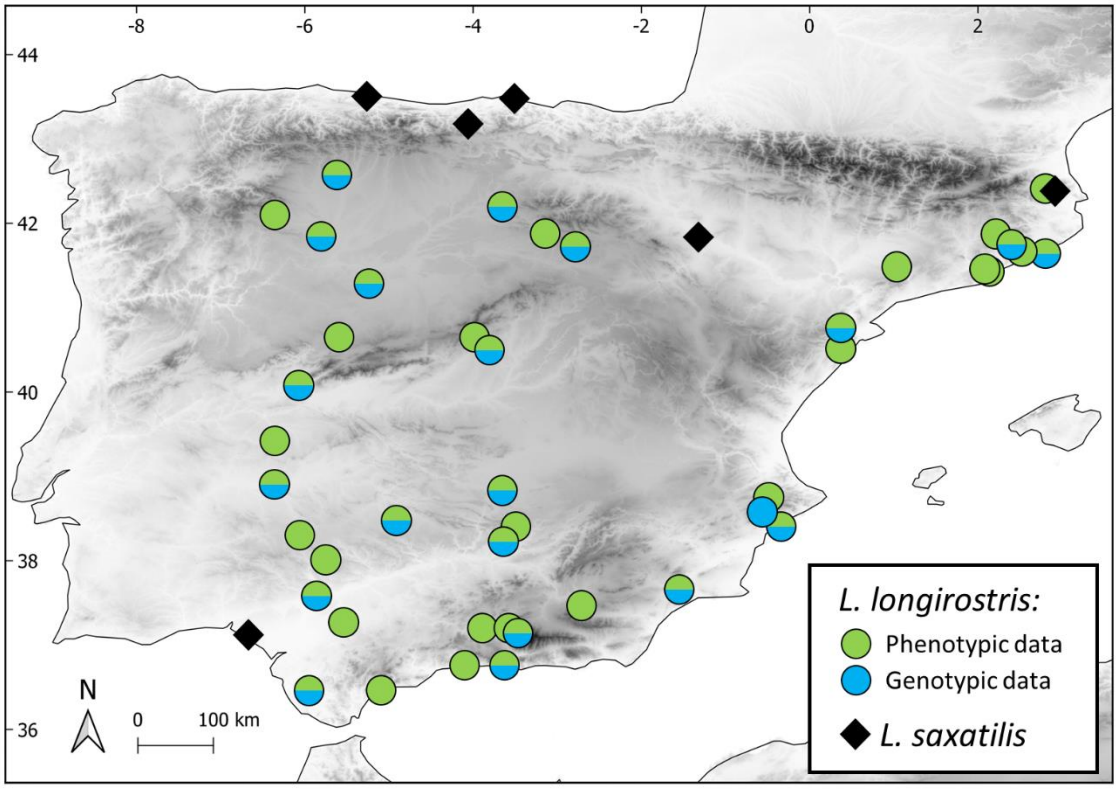


Figure 0.7. Sampled *Leontodon longirostris* and *L. saxatilis* populations. For *L. longirostris*, populations with phenotypic and/or genotypic data used in the experiments are indicated.

In **Chapter 1**, population structure was inferred using principal component analysis (PCA) and clustering methods (FASTSTRUCTURE). To infer the demographic history of the species, two complementary methods were used based on site frequency spectrum (SFS) and likelihood approximations at the population level: one method was not constrained by a predefined demographic model (STAIRWAY PLOT), while the alternative method was carried out using predefined scenarios to get the best fitted model matching the observed data (FASTSIMCOAL2). Once demography was inferred, we compared genetic diversity, as well as additive and recessive genetic load, in three groups of populations distributed along the expansion route (*core*, *intermediate* and *front*).

In **Chapter 2**, we performed two common garden experiments (Fig. 0.8A and B). In each experiment we used a variable number (15-26) of families from 42 *L. longirostris* populations and 5-6 *L. saxatilis* populations. In each experiment, more than 1,100 seeds were sown following a randomized complete block design. We measured

the following traits: germination rates in early fall and late winter conditions; germination time (days from seeding to the emergence of cotyledons); flowering time (days from germination to the emergence of the first flowering stalk, Fig. 0.8C and D); senescence time (days from flowering to death or removal at the end of the experiment); size at flowering (total rosette area), reproductive biomass (Fig. 0.8E), vegetative biomass, below-ground biomass (Fig. 0.8F) and reproductive vs. vegetative allocation. We then assessed phenotypic correlations among traits, as well as the direct and indirect sequential effects of phenology on final plant size as a proxy for fitness (path analysis). Trait variation was also correlated with several climatic variables to ascertain the potential selection agent related to the observed phenotypic variation.

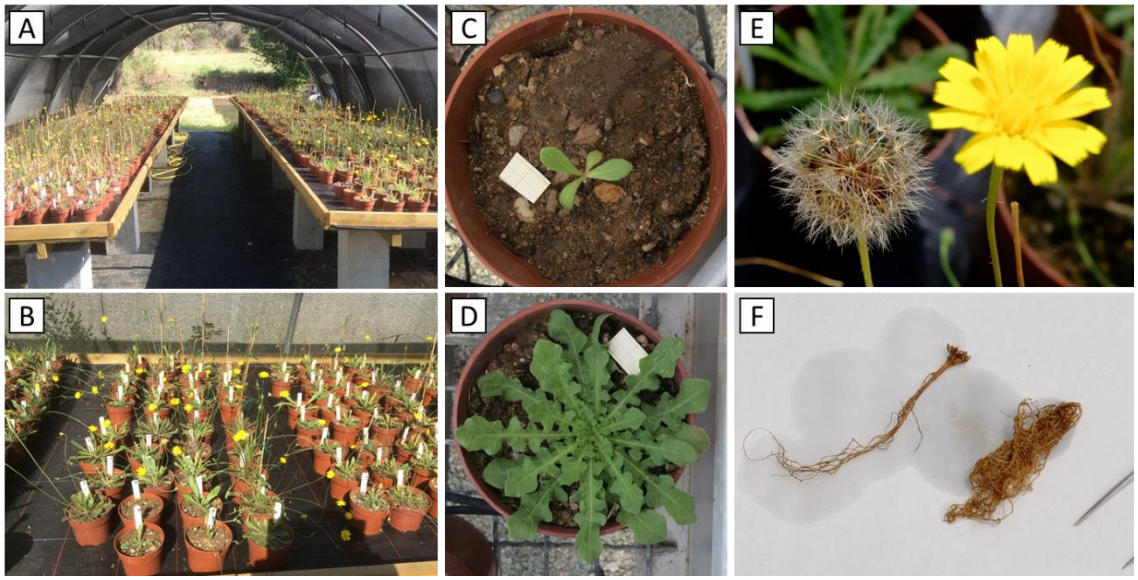


Figure 0.8. Common garden experiment and different developmental phases and features of individual plants. **A**) Common garden settings in a randomized complete block design; **B**) Detail of some pots; **C** and **D**) Plants bolting (onset of the first flower stalk); **E**) Flower and capitulum with seeds; **F**) Cleaned roots before weighing.

Finally, in **Chapter 3**, we used the genomic data to investigate signatures of selection at molecular level in *L. longirostris* populations along the expanding route. We used several complementary approaches. First, we applied two methods to detect signatures of hard (SWEED) and soft sweeps (G12) in different regions of the species' range (*core* and *front*). Second, we used two association methods (BAYESCENV and BAYPASS) that take into account population genetic structure to search for significant associations between the SNPs (single nucleotide polymorphisms) and phenotypic traits

(genotype-phenotype associations, GPA), as well as with specific climatic variables (genotype-environment associations, GEA).

Chapter 1

Demography, genetic diversity and expansion load

This chapter has been published to *Molecular Ecology* (2021), DOI:10.1111/mec.15802.

1.1 Introduction

Population expansions and range shifts are common processes in response to environmental change, often involving complex interactions between demographic and genetic processes. Thus, they have been extensively studied in ecology and evolution (see Chuang & Peterson, 2016 for a review). Theoretical works have demonstrated that the demographic processes taking place during range expansions translate into particular footprints at the genomic level (Excoffier et al., 2009). On the one hand, the strong and consecutive bottlenecks create gradients of decreasing diversity along the expanding range (Austerlitz et al., 1997). On the other hand, the low population densities and strong genetic drift at the expansion front can promote "surfing" of new and standing variants that quickly fix and spread over large regions (Edmonds et al., 2004; Klopfstein et al., 2006). This surfing effect can result in an accumulation of deleterious mutations due to inefficient selection, leading to a steady decrease in mean fitness at the expansion front, the so-called "expansion load" (Peischl et al., 2013).

Over the past decades, many phylogeographic studies in plants and animals from the Northern Hemisphere have supported some of these predictions. For example, an overall decline in genetic diversity is often found along postglacial colonization routes (Hewitt, 2000, 2004). Direct experimental evidence for the accumulation of deleterious mutations following range expansions is, however, very limited (e.g., Bosshard et al., 2017 in bacteria), and indirect empirical evidence mainly comes from a number of studies of the out-of-Africa expansion of human populations (e.g., Fu et al., 2014; Henn et al., 2016; Peischl et al., 2016). In plants, studies supporting expansion load are restricted to species with substantial genomic resources such as the model tree *Populus trichocarpa* (Zhang et al., 2016), the annual herb *Mercurialis annua* (González-Martínez et al., 2017), and three Arabidopsis satellite systems (all Brassicaceae), i.e. *Arabis alpina* (Laenen et al., 2018), *Eutrema salsugineum* (Wang et al., 2018) and *Arabidopsis lyrata* (Willi et al., 2018). Because both low genetic diversity and a high frequency of deleterious alleles can constrain the ability of populations to adapt to novel environmental conditions, it is thus relevant in the context of current climate change to investigate the prevalence of genetic surfing and expansion load in other plants with distinct life-history traits.

Theoretical models have shown that deleterious alleles can persist in range-front populations for thousands of generations after the end of the expansion (Peischl et al.,

2013). Nevertheless, the impact and magnitude of the expansion load depends on a number of factors. In general, expansion load increases under conditions that favor genetic surfing, such as small population size, high growth and low migration rate from neighbor populations (Klopfstein et al., 2006; Peischl et al., 2013). Conversely, mechanisms favoring increased migration to range-front populations throughout the course of range expansion, such as the evolution of higher dispersal rates at the leading edge ("spatial sorting", Shine et al., 2011), could reduce genetic drift and "rescue" populations from incurring expansion load (Peischl & Gilbert, 2020; Tomasini & Peischl, 2020). In the same way, when range expansions occur along environmental gradients, the expansion is slowed by the need for colonizing populations to adapt to the novel local environment. This allows more time for migrants to reach the range-front, restoring genetic diversity, and for natural selection to reduce the frequency of deleterious alleles (Gilbert et al., 2017). Then, species experiencing expansions along heterogeneous habitats and with high dispersal capacity may prevent and/or mitigate the accumulation of deleterious mutations in range-front populations.

Species colonizing disturbed and newly human-made habitats are ideal systems to gain insight into the factors that can enhance or mitigate expansion load. Colonizing plants can spread quickly into ecosystems where they have not been before, so they are expected to experience frequent demographic and spatial expansions. Some attributes of colonizing plants, such as founder events involving small numbers of individuals and/or high growth rates, are expected to favor gene surfing and the accumulation of deleterious alleles. By contrast, colonizing plants display high dispersal capabilities that can mitigate the accumulation of deleterious mutations in range fronts by reshuffling genetic diversity from neighboring populations. Moreover, most colonizing species form temporary seed banks allowing them to survive during unfavorable seasons. By increasing the effective size of populations, seed banks might also play a key role to mitigate genetic drift and load following range expansions. Finally, colonizing species usually expand their range along environmental gradients, which may require quick adaptive responses (e.g., Montague et al., 2008; Colautti & Barrett, 2013). The need to adapt to the novel environmental conditions found in the expanding front can slow the expansion rate and, consequently, the accumulation of deleterious mutations (Gilbert et al., 2017).

Leontodon longirostris (Finch & P.D. Sell) Talavera (≡ *Leontodon saxatilis* subsp. *rothii* Maire ≡ *Thrincia hispida* Roth) (Asteraceae, Cichorieae) is a common colonizing species in the Western Mediterranean Basin that grows in abandoned agricultural fields, therophytic grasslands, roadsides, and other disturbed spaces on a variety of soils (Talavera et al., 2015). It is also widely naturalized in other regions with Mediterranean climate such as Chile, parts of the United States and southern Australia, where is considered to be invasive (CGP; Groves et al., 2003). *Leontodon longirostris* usually behaves as an annual plant, and it is a self-incompatible outcrossing species pollinated by generalist insects (Ruiz de Clavijo, 2001). As some other Asteraceae, it produces dispersible and non-dispersible diaspores, a mixed strategy traditionally interpreted as useful for increasing survival in highly unpredictable habitats (Venable & Lawton, 1980). Central achenes bear well-developed dispersal structures (i.e., a pappus) and exhibit limited dormancy, while peripheral achenes lack a pappus and remain in a dormant state on the ground, forming a temporary seed bank (Ruiz de Clavijo, 2001). This enables the spreading of offspring in space and time, since the central achenes are easily dispersed into new habitats, while the peripheral ones can persist in the established populations until the arrival of favorable conditions for germination.

The species is widely distributed throughout the Iberian Peninsula, a territory characterized by a marked latitudinal gradient of temperature and precipitation. It has one of the smallest genomes within the Asteraceae (Vallès et al., 2013), and a high-quality reference transcriptome assembly is already available for the sister species *L. saxatilis* Lam. (≡ *Thrincia saxatilis* (Lam.) Holub & Moravec, ≡ *Leontodon taraxacoides* (Vill.) Mérat, nom. illeg.) (Hodgins et al., 2014), which greatly facilitates population genomic studies. Then, it provides a particularly good system to gain insight into the factors that can enhance or mitigate expansion load, and thus contribute to the current debate on the potential for species' ranges to shift in response to climate change. With this goal in mind, we generated a genome draft assembly and annotation for *L. longirostris* and used it to design target capture probes to address the following questions: (i) what is the specific demographic history of the species in the Iberian Peninsula?; (ii) if range expansions took place, can we still detect the predicted loss in genetic diversity and increase in genetic load along the expansion axis in this widespread colonizing plant?; and (iii) how the accumulation of deleterious mutations

in front-range populations compares to what is observed in plant species with other life-history traits?

1.2 Material and methods

Sampling for DNA extraction

Leaves from 7 to 20 plants (mean = 12) of *L. longirostris* were collected at 21 localities following three south-to-north transects across the Iberian Peninsula (western, central and eastern, Fig. 1.1, Table S1.1), resulting in a total sample size of 248 individuals. An additional sample from northeastern Iberia was used to generate the genome draft. We also sampled 20 individuals of the close relative *L. saxatilis* from two localities in northern Iberia to be used as outgroup (Fig. 1.1). High quality genomic DNA was isolated from 50 to 100 mg of dry leaf material using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following standard protocols.

Genome draft and annotation

Short-read data are provided on BioProject PRJNA648858 and correspond to 1/2 lane of Illumina HiSeq 2000 (2 x 100 bp) reads from one library with 300- to 500-bp insert size, performed at GATC Biotech, Konstanz, Germany. Raw reads were preprocessed and filtered using SEQTRIMNEXT V2.0.67, a next-generation sequencing-evolved version of SEQTRIM (Falgueras et al., 2010), with default parameters to remove adapters, low-quality base calling, PCR duplicates, short or empty insert sequences, and contaminants (including microorganisms, organelles and plasmids). Useful, pre-processed reads were then connected, when overlapping, using *k*-mer frequencies to conform a long read using COPE V1.1.3 (Connecting Overlapped Pair-End reads; Liu et al., 2012) since average insert size was smaller than the sum of the two read length. Reads were then assembled using three different assemblers, RAY v2.3.1 (Boisvert et al., 2012), SOAPDENOVO2 v2.40 (Luo et al., 2012) and VELVET v1.2.10 (Zerbino & Birney, 2008), with *k*-mers fixed at 31, 43, 57, 71, and a complete range combination from 24 to 71. Scaffolds were obtained from each assembling procedure and submitted to a gap-filling step with GAPCLOSER v1.12 (Li et al., 2010), reusing the useful reads, to provide the final set of contigs and scaffolds with as few indeterminations and gaps as possible.

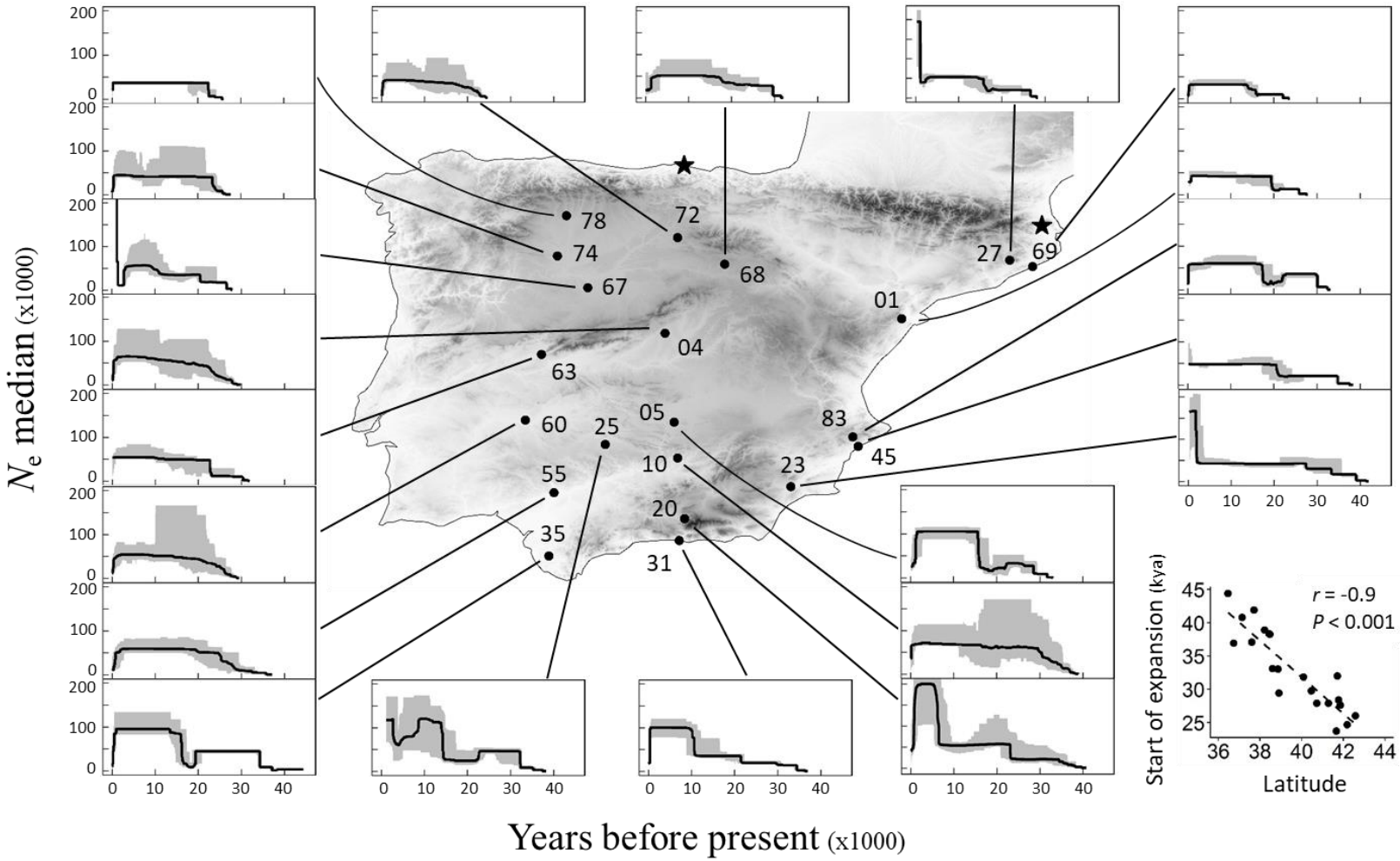


Figure 1.1. Geographic distribution of the populations sampled and inferred demographic history of *Leontodon longirostris* across the Iberian Peninsula. Estimated changes in population size were obtained separately for each population with the STAIRWAY PLOT method using 116,946 SNPs with known allelic state (see text for further details). Black lines in plots show the changes in median effective population size (*N*e) for the last 40,000 years, and shaded areas indicate the 95% confidence intervals. The bottom right figure shows the correlation between the inferred start of expansion (in thousands years before the present) and the geographical position (latitude) of each population. Details on the populations are given in Tables 1.1 and S1.1. The two populations of *Leontodon saxatilis* used as outgroup are indicated by black stars.

Useful reads were then mapped to the resulting contigs and scaffolds using BOWTIE v2.2.9 as a measure of quality and representativity. The assembly produced by SOAPDENOVO2 using the *k*-mer range 24-71 produced a slightly smaller number of contigs and scaffolds than other assemblies (410,019 contigs + scaffolds), but also longest contigs and scaffolds and a highest mapping rate, and therefore it was chosen as the final genome draft for subsequent analyses (Claros et al., 2020; see below).

Gene prediction in the final set of contigs and scaffolds was performed depending on their length. Scaffolds ≥ 10 kbp were annotated using MAKER v2.31.6 (Cantarel et al., 2008) trained with the available transcripts of *L. saxatilis* produced by Hodgins et al. (2014) and the full-length plant proteins found in the UniProtKB database. Gene models were then converted in coding sequences and their tentatively coded protein annotated with FULL-LENGTHERNEXT v0.9.8 (P. Seoane et al., in preparation; available at http://www.scbi.uma.es/ingebiol/commands/full_lengther_next). The remaining contigs and scaffolds, whose length cannot presumably contain a complete gene model, were loaded into our GENEASSEMBLER v0.1.0 pipeline (available at https://rubygems.org/gems/gene_assembler; Seoane-Zonjic et al., 2016) with the aim of joining contigs and/or scaffolds containing sequences from the same gene to generate and annotate fragmented gene models based on the full-length proteins of *Helianthus annuus* found at the UniProtKB database. These "chimaeric" gene models were also annotated with the UniProtKB proteins used as model. Sequences of the 410,019 scaffolds and contigs conforming the *L. longirostris* genome draft, their MAKER-based gene models of the longest (>10 kb) fragments, their functional annotation, and the sequence of the "chimaeric", GENEASSEMBLER-based gene models reconstructed from short (<10 kb) genome draft fragments are publicly available with DOI: 10.6084/m9.figshare.12706247.v3.

Probe design, library construction, and targeted sequencing

We used the annotated draft genome to select ~1.59 Mbp for targeted sequencing in 248 individuals of *L. longirostris* and 20 of *L. saxatilis*. The detailed selection procedure was the following. We first focused on the 536 gene models with a functional ortholog in the protein databases that were located on the longest scaffolds (lengths ≥ 10 kbp; see Results), where more accurate gene predictions and functional annotations were

expected. We refined this dataset discarding those scaffolds containing duplicated annotations and putative gene models with uncharacterized functions. For the remaining gene models, we assessed their biological functions by using the information available at UniProt, and prioritized those involved in processes that can result in adaptive variation (e.g., resistance to cold and drought, resistance to pathogens, phenological traits such as germination and flowering, etc.) over those related to biological functions common to distinct organisms (e.g., DNA replication). This resulted in a dataset of 93 scaffolds containing 249 distinct gene models. We repeated the same procedure with the 472 annotated gene models located on scaffolds with shorter lengths (< 10 kbp; see Results), and retained 50 additional gene models, so the final dataset to be investigated in the whole population sample consisted of 299 gene models located on 143 contigs or scaffolds with a total length of ~1.59 Mbp. Approximately one third of the 1.59 Mbp selected corresponded to genic (coding and non-coding) regions (0.58 Mbp), while the remaining two thirds were intergenic (upstream and downstream) regions (1.01 Mbp).

The selected contigs and scaffolds were sequenced in the 268 individuals sampled (including outgroups) by a targeted sequence capture approach using a custom SeqCap EZ design (Roche NimbleGen Inc, Basel, Switzerland) followed by next-generation sequencing of captured regions on an Illumina platform. Probes were designed by Roche Tech-Support in Madison (WI, USA) starting from the selected target sequences. Library preparation and targeted sequencing were outsourced to IGA Technology Services (Udine, Italy). Libraries were constructed by using KAPA DNA Library Preparation Kit (Roche, NimbleGen Inc.) following the manufacturer's protocol and enrichment performed using NimbleGen solution-based SeqCap EZ probe libraries kit. Cluster generation, template hybridization, isothermal amplification, linearization, blocking and denaturization, and hybridization of the sequencing primers were then performed on Illumina cBot and flow cell HiSeq SBS V4 50 cycle kit, loaded on HiSeq 2500 Illumina sequencer producing 50 bp single reads. The CASAVA version of the Illumina Pipeline 1.8.2 was used for base calling and demultiplexing. Adapters were masked using CUTADAPT v3.0 (Martin, 2011). Masked and low-quality bases were filtered using ERNE-FILTER v2.1.1 (Del Fabbro et al., 2013).

The Genome Analysis Toolkit GATK was used for SNP calling, following GATK best practices for SNP filtering and the following filtering expression: "MQ0> = 4 && ((MQ0 / (1.0 * DP)) > 0.1) || DP < 10.0 || Q < 50.0 || QD <1.5 || FS > 60". SNP

data were further filtered using VCFTools v0.1.15 (Danecek et al., 2011) and vcffilter (vcflib C++ library) for mean coverage across all samples between 20 and 250, a maximum level of missing data per SNP of 10% and a maximum level of missing information per individual of 25%. Ten individuals of *L. longirostris* were discarded with this filter, leaving a final sample size of 238 individuals from 21 populations (Table 1.1). The dataset was additionally filtered removing those SNPs with a significant heterozygote excess in ≥ 3 populations to minimize the impact of putative paralogous loci on data analysis. Finally, only biallelic SNPs were retained, leaving a final dataset of 168,733 polymorphic SNPs distributed along intergenic regions (63%), exons (13%) and non-coding sections of genes (24%; Table S1.2).

Table 1.1. Sample size (*N*) and genetic diversity statistics for the populations of *Leontodon longirostris* included in this study. The number of polymorphic sites, F_{IS} and Tajima's *D* (Tajima, 1989) were based on the 116,946 SNPs with known allelic state. Nucleotide diversity θ_π (Tajima, 1983) per site × 10^{-3} was obtained for synonymous ($\theta_{\pi S}$) and non-synonymous ($\theta_{\pi N}$) sites using a subset of 14,381 SNPs found in exons. Names and geographical location of the populations are provided in Table S1.1. Details for *core*, *intermediate* and *front* groups of populations are provided in the main text.

Population code	***N***	**Number of polymorphic sites**	F_{IS}	**Tajima's *D***	$\theta_{\pi N}$	$\theta_{\pi S}$	$\theta_{\pi N}/\theta_{\pi S}$
10	1	28,440	-0.027	-1.511	1.64	6.85	0.239
20	2	41,957	0.015	-1.765	1.53	6.67	0.229
23	1	26,124	0.019	-1.233	1.47	6.49	0.227
25	7	22,235	-0.003	-1.444	1.42	6.17	0.230
31	1	30,422	0.069*	-1.550	1.35	6.49	0.208
35	1	27,782	0.028	-1.320	1.43	6.96	0.206
45	1	27,323	0.030	-1.297	1.30	5.77	0.225
55	1	22,635	-0.002	-1.189	1.25	6.00	0.209
Core	9	88,098	-	-2.110	1.47	6.55	0.222 ± 0.012
04	9	20,790	0.070	-1.494	1.23	5.64	0.219
05	9	23,307	0.041	-1.542	1.22	5.66	0.216
60	9	19,773	0.060	-1.288	1.28	5.28	0.242
63	9	21,371	0.044	-1.464	1.19	5.01	0.237
68	1	20,295	-0.072	-1.263	1.12	5.95	0.189
83	1	25,995	0.042	-1.414	1.24	5.67	0.219
Intermediate	5	60,400	-	-2.084	1.24	5.56	0.220 ± 0.019
01	6	14,383	0.115*	-0.921	1.30	6.29	0.207
27	1	25,194	0.041	-1.733	1.14	5.00	0.229
67	1	23,002	-0.006	-1.393	1.11	5.00	0.223
69	1	17,373	0.039	-1.293	1.02	4.10	0.247
72	1	16,402	0.054	-1.373	1.01	4.29	0.235
74	1	18,229	0.000	-1.320	1.10	4.92	0.224
78	1	17,795	0.053	-1.386	1.05	4.60	0.228
Front	8	52,330	-	-2.028	1.10	4.81	0.228 ± 0.012
Overall	2	116,946	-	-2.249	1.32	5.83	0.226

* $P < 0.01$

Genetic diversity and population structure

We applied two different approaches to infer population genetic structure in *L. longirostris* at the regional level using the full SNP dataset (168,733 SNPs). First, the existence of discrete clusters was explored using the Bayesian clustering method implemented in FASTSTRUCTURE v1.0.4 (Raj et al., 2014). Three independent runs for each K were performed, from $K = 1$ (no structure) to $K = 10$, and averaged Q values (i.e., the individual assignment probability to each of the K groups) were used to draw pie charts using QGis v2.14.0-Essen (Quantum GIS Development Team, 2016). Second, given that model-based clustering methods tend to overestimate the number of discrete clusters when genetic variation is continuously distributed across the landscape (e.g., under isolation by distance; Meirmans, 2012), a principal component analysis (PCA) was performed using PLINK v2.00a (Chang et al., 2015) with default parameters.

For subsequent analyses (i.e., population genetic diversity and differentiation, neutrality tests and demographic inference), we refined the SNP dataset and used only those SNPs for which the state of the allele (ancestral vs. derived) could be inferred by comparison with *L. saxatilis* (116,946 SNPs). Nucleotide diversity θ_π (Tajima, 1983) and Tajima's D neutrality test (Tajima, 1989) were computed using MSTATSPOP (https://bioinformatics.cragenomica.es/numgenomics/people/sebas/software/software.html) on concatenated sequences, both for each population and considering all populations as a whole. In addition, the efficacy of selection evaluated as the ratio of non-synonymous to synonymous nucleotide diversity ($\theta_{\pi N}/\theta_{\pi S}$) was computed based on the 14,381 SNPs from coding regions. The percentage of heterozygous sites was calculated at the population level with VCFTools v0.1.15 (Danecek et al., 2011). Since departures from random mating could be indicative of restricted migration and/or changes in the mating system, two processes that are expected to decrease genetic diversity and increase expansion load, average inbreeding coefficients within (F_{IS}) and between (F_{ST}) populations were obtained using ARLEQUIN v3.5.2.2 (Excoffier & Lischer, 2010) and their significance was evaluated with 10,100 permutations. Isolation-by-distance (IBD) was evaluated according to Rousset (1997) by testing the correlation between the matrix of pairwise [$F_{ST}/(1- F_{ST})$] and the matrix of geographical distances (logarithmic scale) using the Mantel test implemented in ARLEQUIN v3.5.2.2, with 10,100 permutations. Finally, the relationship between genetic diversity with latitude and longitude was tested using Pearson correlations in R v3.4.4 (R Core Team, 2019).

Demographic history

We used two complementary approaches based on the unfolded site frequency spectrum (SFS) to infer the most likely demography of *L. longirostris*. On the one hand, the model-flexible approach implemented in STAIRWAY PLOT v2 (Liu & Fu, 2015) was applied to infer the demographic history of each population separately. This method is not restricted to a specific demographic model, so it can infer significantly more detailed demographic history than model-constrained methods, being more suitable for demographic analysis where no previous knowledge is available. We assumed a per-generation mutation rate of 6.5×10^{-8} per base pair, that was the mutation rate inferred considering a demographically stable population ("standard model") in FASTSIMCOAL2 (see below), and a generation time of one year (as the species usually behaves as an annual herb). Median estimates of the effective population size (N_e) and confidence intervals were estimated with the built-in bootstrap function using 200 subsets of the input data.

However, the method implemented in STAIRWAY PLOT does not allow testing the goodness-of-fit of the expected SFS with the observed SFS. Then, we conducted a second complementary approach to fit specific demographic models by maximum likelihood to the unfolded SFS using FASTSIMCOAL2 v2.6.0.3 (Excoffier et al., 2013). Since similar and simple demographic histories were inferred with STAIRWAY PLOT for all populations, four basic demographic models of population change were compared (Fig. 1.2). In the first three, the parameters were restricted to fit particular demographic scenarios: (1) a "standard model" for a demographically stable population, defined by a single parameter, the current population size; (2) a "bottleneck + expansion model", characterized by a reduction in population size followed by a population growth; and (3) an "expansion + bottleneck model", with a period of population growth followed by a decrease in population effective size. A fourth 3-epoch model (4) with no assumptions about the past demographic events (i.e., no parameter restriction) was also included. This model consisted of three effective population sizes allowed to change at two different times in the past (Fig. 1.2). As in the STAIRWAY PLOT approximation, the mutation rate was inferred from the "standard model" (6.5×10^{-8}) and the generation time was set to one year.

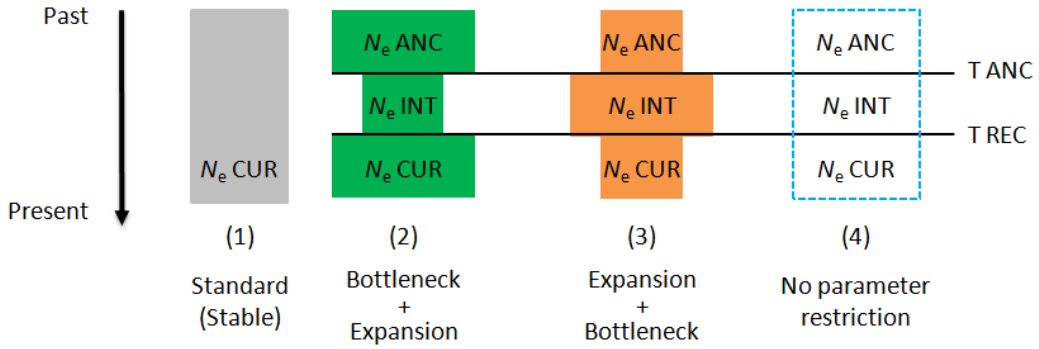


Figure 1.2. Schematic representation of the four alternative demographic models tested using FASTSIMCOAL2 (see Table S1.3 for details on the search ranges of the parameters). Note that for the model (4) there were no restrictions regarding the direction of the population size change, as it allowed for either increases or decreases in population size. *Ne* ANC: ancestral effective population size; *Ne* INT: effective population size between ancestral and cu rrent periods; *Ne* CUR: current effective population size; T ANC: ancestral time of population size change; T REC: recent time of population size change.

Each model was run 100 replicated times, with 100,000 coalescent simulations for the calculation of the composite likelihood and 20 expectation-conditional maximization (ECM) cycles (see Table S1.3 for search ranges). The run with the maximum likelihood in each scenario was retained and used to obtain point estimates of the different demographic parameters. The best runs in each model were then compared using AIC and the relative likelihood (Akaike's weight of evidence) to select the most probable demographic scenario, as in Excoffier et al. (2013). Confidence intervals (95%) were obtained for each parameter of the best scenario using parametric bootstrap replicates. One hundred SFS were simulated from the run with the maximum composite likelihood, and then parameters were re-estimated performing 20 runs per simulated SFS. The runs with the highest maximum likelihood per simulated SFS were then used to define the 95% CI values (Excoffier et al., 2013).

Finally, to infer the possible origin and direction of a range expansion, we applied the method of Peter & Slatkin (2013) to detect asymmetries in the 2D site frequency spectrum between pairs of populations. The directionality index (ψ) was calculated using the R package rangeExpansion v.0.0.0.9000 (freely available at https://github.com/BenjaminPeter/rangeexpansion).

Expansion load

To evaluate the potential effect of range expansions on genetic load, the 116,946 SNPs with known allelic state (ancestral vs. derived) were classified into four categories according to the predicted effect of the variant change, using SNPEFF v4.3t (Cingolani et al., 2012) and the genome draft as a reference: HIGH (e.g., loss of start and stop codons, 299 SNPs), MODERATE (e.g., non-synonymous mutations, 7,562 SNPs), LOW (e.g., synonymous mutations, 7,642 SNPs) and MODIFIER (e.g., non-coding variants or variants affecting non-coding parts of genes, 101,443 SNPs). We then retained the SNPs included in the first three categories (15,503 SNPs) for the estimation of genetic load, discarding the SNPs located in non-coding regions where the impact of the variant change is difficult to predict. About 96% of these SNPs were located on the largest scaffolds (≥ 10 kbp), where predictions of gene models were expected to be more accurate.

We used several proxies to evaluate the impact of the expansion on the spread and fixation of deleterious alleles. First, we counted the total number of fixed derived mutations at the population level for each SNPEFF category (HIGH, MODERATE and LOW). Second, we computed the additive and recessive genetic load at the individual level and tested for differences between three different groups of populations. The additive genetic load, defined as the number of potentially deleterious derived alleles per individual, was obtained by counting the derived mutations classified as HIGH and MODERATE in homozygous (x2) and heterozygous (x1) state. The recessive genetic load was obtained as the number of potentially deleterious derived alleles (i.e., HIGH and MODERATE) in homozygous state. Both measures were scaled by derived mutations with LOW effect (which are likely to be neutral or just slightly deleterious) in order to evaluate differences in the efficacy of purifying selection at individual level.

The groups of populations were defined using the information obtained in the demographic analyses, which suggested that *L. longirostris* progressively expanded its range from south to north in the Iberian Peninsula (see Results for further details). To determine whether this range expansion has resulted in a higher genetic load due to relaxed selection in range-front populations (i.e., expansion load), we compared the additive and recessive genetic load among three groups of populations according to their geographical position and the starting time of the demographic expansion inferred with STAIRWAY PLOT (see Results). The first group (*core*) included those populations from

southern latitudes that have been estimated to be present for more than 35,000 yrs BP (population's ID: 10, 20, 23, 25, 31, 35, 45, 55). The second (*intermediate*) was composed by populations at intermediate latitudes that were estimated to colonize its current range during the past 29,000-33,000 yrs (population's ID: 04, 05, 60, 63, 68, 83). Finally, the third group (*front*) comprised those populations that reached northern latitudes not before 28,000 yrs ago (population's ID: 01, 27, 67, 69, 72, 74, 78). Averaged values of recessive and additive genetic load were computed for each group and significant differences among them were evaluated using Mann-Whitney non-parametric tests. In addition, to distinguish the effect of range expansion on more ancient (pre-existing) variation vs. new (more recent) mutations, we compared the averaged additive and recessive genetic load obtained for sites with derived mutations shared between *core*, *intermediate* and *front* groups of populations (3,910 SNPs) with those obtained for private sites within each group (5,015 SNPs in the *core*, 1,489 SNPs in the *intermediate* and 1,565 SNPs in the *front*).

1.3 Results

Leontodon longirostris genome draft

From the original 186,552,004 paired-end reads, 13,701,592 were discarded, mainly due to their chloroplastic origin. Most remaining reads were still paired-end (91.13% or 170,006,980 reads), indicating that they were suitable for genome assembling. After 20 assembling approaches using different assemblers, *k*-mers and consolidation strategies, SOAPDENOVO2 using the *k*-mer range 24-71 presented the best compromise since it had a high mapping rate (82.14%) and produced a reasonable amount of scaffolds and contigs (Table 1.2). It provided an assembly length of 418 Mb covering ~53% of the genome (total size: 0.78 Gb, Vallès et al., 2013). Low N50 (1,532 bp) indicated a fragmented genome draft, which was an expected result due to the absence of long paired-end reads among the input read for assembling. Gene predictions were obtained with different bioinformatic approaches for the 1,007 longest scaffolds and the remaining "short" fragments of the genome draft (see Material and methods). From long scaffolds, 853 gene models were inferred suing 536 different proteins; 202 of the models contained the complete ORF; additionally, other 169 gene predictions coded for an unknown protein (Table 1.2). Regarding the small fragments, further 472

"chimaeric" gene models were reconstructed (Table 1.2), only five using the same protein than those from the longest scaffolds (P45739, Q8LSM7, E3SU13, F8R6K3 and P85200). Therefore, at least 1,003 different genes were identified in the *L. longirostris* genome draft (DOI: 10.6084/m9.figshare.12706247.v3) based on known proteins, a number that can be extended to 1,172 when the 169 putative genes without ortholog are also considered. A brief consultation in PANTHER (DOI: 10.1038/s41596-019-0128-8), where only 471 orthologs were recognized, indicated that the most important groups of functions corresponded to "catalytic activity" (GO:0003824) and "metabolic process" (GO:0008152), including at least 113 and 103 of the predicted genes, respectively.

Table 1.2. Summary of the assembly features and the gene prediction for the *Leontodon longirostris* draft genome (DOI: 10.6084/m9.figshare.12706247.v2).

	Values or counts
Draft genome features	
Number of scaffolds + contigs	410,019
Scaffolds > 10 kb	1,007
Longest scaffold (bp)	53,452
N50 (bp)	1,532
Total length (Mbp)	418
Mapping rate (%)	82.14
Gene predictions in the genome draft	
In long scaffolds	853
Unique orthologs	536
Complete unique orthologs	202
Putative coding without ortholog	169
In short scaffolds + contigs	472
Common model IDs in short and long scaffolds	5
Total different gene models	1,172

Genetic diversity and population structure

Based on the targeted sequencing data, all the partitions in FASTSTRUCTURE (K = 2-10) supported a clear separation between southern and northern populations, with populations of admixed composition at the intersection of both groups (Fig. S1.1). However, no discrete groups were found in the PCA, where the distribution of

individual genotypes along the first axis (PCA1; 25.3% of the variation) mostly reflected a south-to-north latitudinal cline (Fig. 1.3). A longitudinal separation of northernmost populations was also evident in the PCA2 axis (10.6% of the variation). The third and fourth axes (9.1% and 8.8% of the variation, respectively) revealed the separation of some particular populations (e.g., 35, 78) or groups of geographically close populations (e.g., 23, 45, 83; Fig. S1.2).

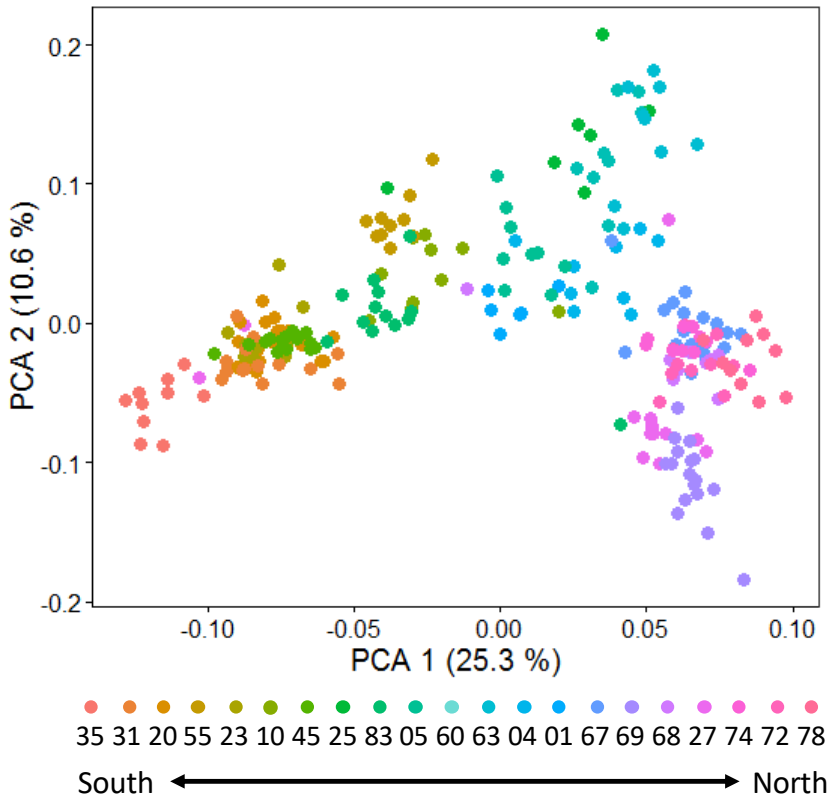


Figure 1.3. Principal component analysis (PCA) of the genetic variation found in 238 individuals of *Leontodon longirostris* in the Iberian Peninsula (based on 168,733 SNPs). Each individual is represented by a colored point. Note that the distribution of individuals along the PCA 1 axis mostly reflects their latitudinal position from south to north. Population details can be found in Tables 1 and S1.

At the population level, genetic diversity estimates based on the refined SNP dataset with known allelic state (116,946 SNPs) revealed a significant negative correlation with latitude, both for the number of heterozygous sites (r = -0.79, P < 0.001; Fig. 1.4A) and the overall nucleotide diversity (r = -0.92, P < 0.001; Fig. 1.4B). The number of polymorphic loci, as well as the nucleotide diversity present at

synonymous and non-synonymous sites, were also reduced in northern (*front*) populations as compared to the southern (*core*) ones (Table 1.1). The ratio of non-synonymous to synonymous nucleotide diversity ($\theta_{\pi N}/\theta_{\pi S}$), however, did not show any geographic trend, with mean values being very similar for all groups of populations (*core*: 0.222 ± 0.012; *intermediate*: 0.220 ± 0.019; *front*: 0.228 ± 0.012; Table 1.1). Only two populations departed from Hardy-Weinberg equilibrium, supporting random mating within populations (F_{IS} range: -0.072 − 0.115; Table 1.1). Among populations, we found low to moderate pairwise F_{ST}, but most pairs were significantly greater than zero ($P < 0.001$; Table S1.4). Pairwise [$F_{ST}/(1- F_{ST})$] increased with the logarithm of the geographical distance, supporting an isolation-by-distance pattern ($r = 0.56$, $P < 0.001$; Fig. S1.3).

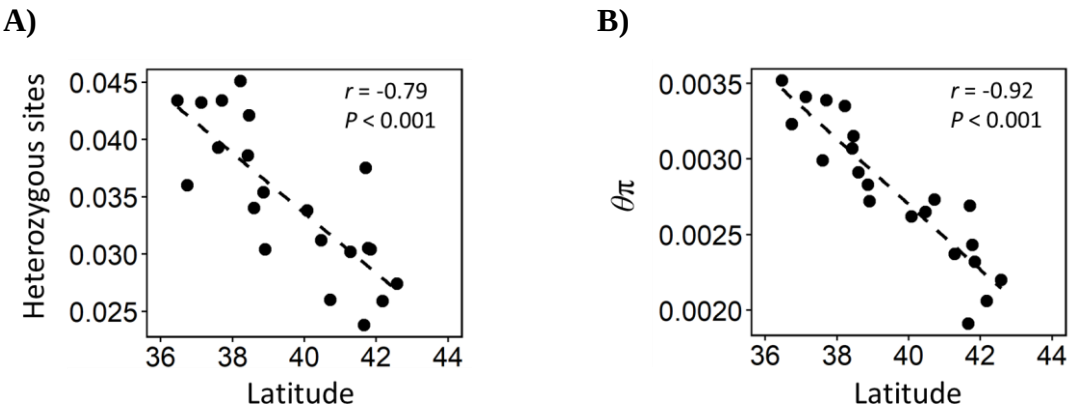


Figure 1.4. Population genetic diversity (based on 116,946 SNPs with known allelic state) in relation to latitude. A) Proportion of heterozygous sites; B) Overall nucleotide diversity θ_π (Tajima, 1983). Pearson coefficients and their significance are reported.

Demographic history

STAIRWAY PLOT modeling revealed that all populations were characterized by demographic expansions, as expected in a colonizing plant. Demographic expansions were also suggested by the negative Tajima's *D* obtained in all populations (Table 1.1). Nevertheless, the onset of the expansion varied among populations following a latitudinal gradient ($r = -0.90$, $P < 0.001$; Fig. 1.1), suggesting a northward spatial expansion of the species across the Iberian Peninsula. The oldest expansions were inferred for southernmost populations and took place around 40,000 yrs BP, reaching mid and high latitudes of Iberia around 30,000 and 25,000 yrs BP, respectively (Fig.

1.1). The directionality index ψ also pointed to a range expansion towards the north, with the most likely origin located in southern Iberian Peninsula (population 35, $P < 0.001$, Fig. S1.4). After a phase of growth following colonization, most populations maintained large and constant effective sizes until recently (N_e : 50,000-100,000), although some of them experienced a demographic bottleneck around 20,000 yrs ago (05, 25, 35, 83). During the last 2,000 yrs, however, a strong demographic decline was evident in most populations (15 out of 21), reaching in some cases current effective sizes as small as 5,000 (Fig. 1.1). In contrast, three populations (23, 27, 67) seem to have experienced a recent exponential growth resulting in much higher current N_e (>150,000).

Statistical comparisons of the four demographic models tested in FASTSIMCOAL2 strongly supported the 3-epoch model with no parameter restriction (Table S1.5). In each population the inferred parameters were generally in accordance with the events inferred with STAIRWAY PLOT (Tables 1.3 and S1.5). The best model for 13 populations (01, 05, 10, 20, 23, 25, 27, 31, 35, 45, 63, 67, 83) consisted of two consecutive demographic expansions, while the remaining eight (04, 55, 60, 68, 69, 72, 74, 78) showed a demographic expansion followed by a decrease in population size occurring recently (~50-2,700 yrs BP), as already detected with STAIRWAY PLOT. In some cases (e.g., populations 05, 20, 31, 35), however, FASTSIMCOAL2 failed to detect the recent demographic decline that was evident with STAIRWAY PLOT, probably because of the simplicity of the models (restricted to 3-epochs). Regarding time estimates, more recent demographic events were inferred with FASTSIMCOAL2 compared to STAIRWAY PLOT, but the oldest expansions were still detected in the *core* (~25,000-32,000 yrs BP), being progressively more recent for *intermediate* (~22,000-26,000 yrs BP) and *front* (~17,000-22,000 yrs BP) populations.

Table 1.3. Parameters inferred with FASTSIMCOAL2 under the best supported demographic scenario in each population. Confidence intervals (95%) are within brackets. White rows indicate populations showing two consecutive demographic expansions. Grey rows indicate populations showing a demographic expansion followed by a recent decrease in population size. Times are given in number of generations, which can be translated directly to years in annual *Leontodon longirostris*, ignoring the effects of a possible seedbank. *Ne* ANC: ancestral effective population size; *Ne* INT: effective population size between ancestral and current periods; *Ne* CUR: current effective population size; T ANC: ancestral time of population size change; T REC: recent time of population size change. Each specific parameter is illustrated in Fig. 1.2. Details on the comparison with the remaining demographic models analyzed with FASTSIMCOAL2 are provided in Table S1.5.

Population code	N_e ANC	N_e INT	N_e CUR	T ANC	T REC
Core					
10	**139** (54-7,819)	**4,378** (3,670-55,916)	**50,074** (48,930-52,170)	**29,540** (28,834-33,909)	**27,431** (11,989-28,135)
20	**587** (219-861)	**24,374** (20,172-27,724)	**74,663** (71,780-77,735)	**29,893** (29,003-31,101)	**10,869** (9,280-12,480)
23	**32** (4-255)	**30,659** (27,687-32,110)	**66,684** (47,371-87,335)	**31,787** (31,215-32,362)	**3,209** (2,322-7,692)
25	**127** (9-397)	**24,433** (18,663-30,846)	**64,308** (58,205-79,076)	**28,825** (27,962-29,435)	**10,110** (6,382-13,949)
31	**63** (9-406)	**16,503** (13,433-21,930)	**54,209** (51,982-56,862)	**28,020** (26,933-28,515)	**12,764** (10,281-14,704)
35	**2,371** (1,643-2,769)	**24,699** (15,183-32,925)	**56,414** (53,308-63,466)	**27,144** (25,746-30,105)	**11,010** (6,599-15,197)
45	**40** (8-596)	**18,536** (15,898-25,140)	**37,935** (36,702-39,914)	**29,504** (27,745-29,796)	**12,507** (7,514-14,908)
55	**1,168** (986-1,365)	**42,571** (40,153-54,632)	**9,990** (2,278-24,639)	**24,862** (24,136-25,436)	**626** (91-4,588)
Intermediate					
04	**358** (258-508)	**42,506** (40,534-62,941)	**32,822** (27,539-38,983)	**21,628** (21,306-22,065)	**2,107** (808-13,782)
05	**9** (3-321)	**12,461** (8,977-21,992)	**52,506** (50,459-56,382)	**25,531** (24,369-25,639)	**15,568** (10,570-17,397)
60	**98** (24-185)	**37,711** (36,505-55,403)	**19,717** (13,746-33,705)	**22,269** (21,920-22,649)	**630** (313-11,958)
63	**21** (3-240)	**16,294** (8,782-28,457)	**41,617** (40,208-45,328)	**24,286** (23,596-24,653)	**14,803** (7,664-18,383)
68	**88** (16-186)	**38,790** (35,992-51,671)	**19,654** (10,538-24,016)	**24,383** (23,931-24,740)	**2,705** (816-7,011)
83	**6** (5-186)	**18,908** (14,040-26,503)	**40,772** (39,541-43,491)	**25,426** (24,900-25,811)	**13,070** (8,262-16,201)
Front					
01	**29** (2-270)	**12,112** (4,109-35,358)	**31,251** (29,676-39,736)	**21,183** (20,327-21,491)	**15,677** (1,693-18,783)
27	**68** (4-199)	**22,271** (17,077-25,785)	**47,718** (44,665-51,429)	**21,665** (21,162-22,066)	**7,558** (5,702-10,475)
67	**5** (2-73)	**26,198** (21,342-26,688)	**47,053** (27,690-62,429)	**21,530** (21,014-21,673)	**413** (274-11,827)
69	**411** (304-697)	**22,089** (21,618-46,247)	**8,334** (5,661-16,835)	**17,091** (16,349-17,491)	**386** (190-6,834)
72	**72** (7-178)	**29,059** (27,667-37,830)	**5,738** (3,723-22,244)	**18,974** (18,458-19,237)	**186** (109-6,351)
74	**350** (232-490)	**34,394** (33,320-42,841)	**1,384** (1,133-21,568)	**20,225** (19,738-20,490)	**51** (41- 3,628)
78	**107** (29-220)	**29,809** (28,784-57,489)	**18,529** (12,449-27,911)	**19,734** (19,337-20,068)	**471** (304-13,664)

Expansion load

The number of derived mutations that were fixed at the population level was extremely low in the three categories considered (i.e., HIGH, MODERATE and LOW). Specifically, we found three derived mutations with MODERATE effect that were fixed in population 35 (*core*), and another three with LOW effect, each one fixed in populations 01 (*front*), 04 (*intermediate*) and 23 (*core*), respectively.

The comparisons among groups of populations revealed that the average number of potentially deleterious mutations (i.e., with HIGH and/or MODERATE effect) carried by each individual decreased along the expansion route (Table 1.4), both for the sites shared between the *core*, *intermediate* and *front* populations (presumably pre-existing variation), and for those that were private (presumably more recent mutations). Nevertheless, once scaled by neutral variation (i.e., variants with LOW effect), additive and recessive genetic load were invariably higher in populations from the *front* as compared to *core* and *intermediate*, suggesting that purifying selection is relaxed in the expansion front (Table 1.4). For shared mutations, the values of additive genetic load increased from 0.52 ± 0.06 and 0.53 ± 0.07 in *core* and *intermediate* populations, respectively, to 0.60 ± 0.09 in those of the *front* (Table 1.4). A similar pattern was observed for shared mutations found in the homozygous state (recessive genetic load), with significantly higher values in the group of *front* populations (0.55 ± 0.37) compared to the rest (0.39 ± 0.15 and 0.37 ± 0.16 for *core* and *intermediate*, respectively). Additive genetic load was also higher for mutations only found in the *front* (1.63 ± 0.79) compared to those only found in the *intermediate* (1.35 ± 0.63) and the *core* (1.16 ± 0.31) populations. Recessive genetic load for private sites was not calculated due to the small number of mutations. It must be noted, however, that standard deviations of genetic load were particularly high in the *front* (Table 1.4), indicating a non-equal contribution of all individuals to the observed values.

Table 1.4. Scaled additive and recessive genetic load (D/ND) per individual in *core*, *intermediate* and *front* populations. Mean and standard deviation (SD) values were obtained for the full number of sites with derived mutations where it was possible to predict the impact of a variant change (All; 15,503 SNPs), for sites with derived mutations shared among the three groups (Shared; 3,910 SNPs), and for sites with private derived mutations within each group (Private; 5,015 SNPs in the *core*, 1,489 SNPs in the *intermediate* and 1,565 SNPs in the *front*). Significant differences among groups ($p < 0.05$) were evaluated using Mann-Whitney non-parametric tests and are indicated with different letters. The scaled recessive genetic load for sites with derived private mutations was not calculated due to small number of mutations. Deleterious mutations (D) include variants classified within the HIGH and MODERATE categories using SNPEFF, while non-deleterious mutations (ND) include those within the LOW one (see text for further details).

	Additive genetic load			**Recessive genetic load**		
Group	**Deleterious mutations (D)**	**Non-deleterious mutations (ND)**	**D/ND**	**Deleterious mutations (D)**	**Non-deleterious mutations (ND)**	**D/ND**
All						
Core	303.87 ± 43.80	508.42 ± 68.45	0.60 ± 0.06 a	47.43 ± 23.87	117.77 ± 57.77	0.42 ± 0.16 a
Intermediate	244.08 ± 49.70	407.97 ± 82.95	0.60 ± 0.07 a	31.29 ± 16.47	78.27 ± 37.29	0.40 ± 0.17 a
Front	220.80 ± 50.86	339.25 ± 81.19	0.66 ± 0.09 b	33.98 ± 18.94	66.02 ± 43.67	0.60 ± 0.34 b
Shared						
Core	203.95 ± 29.72	396.48 ± 49.91	0.52 ± 0.06 a	40.11 ± 19.42	108.74 ± 52.10	0.39 ± 0.15 a
Intermediate	184.17 ± 37.20	349.37 ± 73.12	0.53 ± 0.07 a	27.36 ± 15.07	73.86 ± 35.37	0.37 ± 0.16 a
Front	176.87 ± 41.09	299.90 ± 69.96	0.60 ± 0.09 b	29.29 ± 17.47	62.45 ± 42.20	0.55 ± 0.37 b
Private						
Core	53.20 ± 16.13	48.73 ± 18.46	1.16 ± 0.31 a	3.14 ± 4.24	2.76 ± 4.43	-
Intermediate	22.98 ± 9.87	18.51 ± 8.94	1.35 ± 0.63 a	1.15 ± 1.90	1.19 ± 2.10	-
Front	16.51 ± 7.05	11.81 ± 7.13	1.63 ± 0.79 b	1.14 ± 1.84	0.64 ± 1.53	-

1.4 Discussion

Despite the vast amount of theory dealing with the genetic consequences of range expansions, empirical studies in plants are scarce and most are restricted to model species, limiting generalization. Here, we extended these studies to *L. longirostris*, a non-model Asteraceae that is widely distributed through the Iberian Peninsula. We produced the first genome draft for the species, which provided enough genomic information for subsequently studying nucleotide diversity, demographic history and expansion load via targeted sequencing. We found a northward expansion of the species, which was accompanied by a loss of genetic diversity and an increase in the proportion of putatively deleterious mutations (expansion load), as predicted by theory. Remarkably, signatures of range expansions were still noticeable after thousands of generations. Nevertheless, expansion load did not extend to all the individuals at the range-front, and deleterious mutations were not fixed at the population or the group level, suggesting that several factors could have mitigated the genomic signatures of range expansions in this species, as discussed below.

A first genome draft for Leontodon longirostris

The genome draft for *L. longirostris* obtained in the present study (DOI: 10.6084/m9.figshare.12706247.v3), even if incomplete (a maximum of 1,172 genes were identified), low-covering (only ~53% of the genome was assembled) and fragmented (low N50, Table 1.2), was reliable (read mapping rate of 82.14%) and contained enough and suitable genomic information for further applications. The targeted sequencing approach stemming from this genome draft enabled us to infer the demographic history and expansion load of the species in the Iberian Peninsula, but it could also support population genetic studies oriented to identify candidate genes underlying ecologically important traits in *L. longirostris*, as well as for comparative genomic studies with other *Asteraceae*. Moreover, this draft genome is a base reference that can be further improved, for instance, by the addition of long paired-end reads.

Demographic history of Leontodon longirostris in the Iberian Peninsula

The history *L. longirostris* in the Iberian Peninsula was dominated by both demographic and spatial expansions. Demographical models suggested a progressive expansion of the species' range from south to north, pointing to a possible colonization from northern Africa. The proximity of the Iberian and African land plates, and the sea level drop occurred during the last glacial period, have facilitated intermittent plant migrations across the Strait of Gibraltar, being more frequent in the case of short-lived herbs and pioneer shrubs (Rodríguez-Sánchez et al., 2008; Lavergne et al., 2013). Within the Asteraceae, for instance, *Hypochaeris* sect. *Hypochaeris* and *Helminthotheca*, closely related to *Leontodon*, originated in western North Africa and then expanded through the Strait of Gibraltar to the Iberian Peninsula during the Pleistocene (Ortiz et al., 2009; Tremetsberger et al., 2016). Similarly, a North African origin with subsequent migration into the Iberian Peninsula has been postulated for a relict lineage of the model plant *Arabidopsis thaliana* (e.g., Toledo et al., 2020). Future studies including North African samples of *L. longirostris* could clarify whether the origin of the Iberian samples is located or not in this area.

Contrary to the prevailing pattern of multiple refugia that is commonly found for many plants in Iberia ("refugia-within-refugia" model, Gómez & Lunt, 2007), we did not find divergent lineages within the species. The separation of southern and northern populations inferred by FASTSTRUCTURE is not supported by any other evidence and is probably due to the poor performance of model-based clustering methods in the presence of continuous patterns of genetic differentiation (e.g., "isolation-by-distance"; Frantz et al., 2009). Instead, after the initial colonization of southern Iberia, the species seems to have expanded progressively northward for several millennia. During the Late Pleistocene (125,000-11,500 yrs BP), steppes and grasslands were frequent in the Iberian landscapes (González-Sampériz et al., 2010), as well as large herbivore fauna (Álvarez-Lao & Méndez, 2016). These conditions could have facilitated the spread of *L. longirostris*, allowing at the same time substantial gene flow between different sectors along the expansion axis, precluding the formation of genetically distinct groups. High rates of genetic exchange seem to have occurred over long periods of time, as supported by the low differentiation among neighboring populations and the large effective population size (N_e > 50,000 individuals) inferred for most populations, with only few of them showing signals of demographic bottlenecks matching the Last Glacial

Maximum (around 20,000 yrs BP). Moreover, the gradual increase of forest biomes occurred with deglaciation (14,000 yrs BP) in many parts of Europe (Binney et al., 2017) did not have a significant impact on the size of the populations, which have remained stable until very recent times. During the last two millennia, however, a sharp demographic decline was inferred in some populations (Fig. 1.1, Table 1.3), suggesting a rapid decrease of suitable habitats in much of the species distribution, presumably associated to the agricultural expansion and increasing urbanization.

Genetic consequences of range expansion

A first consequence of the northward range expansion in *L. longirostris* was a remarkable loss of genetic diversity. Despite the relatively small area covered by the expansion (~1,000 km from south to north) and the intrinsic characteristics of the species (e.g., high dispersal ability, outcrossing mating system), the reduction in the number of polymorphic sites (Table 1.1) was comparable to that found in other short-lived plants at much wider geographical scales (e.g., *Mercurialis annua*, González-Martínez et al., 2017). Moreover, while populations of *M. annua* from the expanded ranges exhibited slightly lower values of nucleotide diversity for synonymous and non-synonymous sites than those of the core, the decrease was substantial in *L. longirostris* (Table 1.1). These results could be unexpected given the long time elapsed since the expansion (more than 20,000 generations) and support the idea that the genetic effects of range expansions can be maintained over thousands of generations (Peischl et al., 2013), even in the presence of frequent gene flow. Interestingly, though dispersal seems to have a significant role to attenuate the negative effects of gene surfing on the fixation of deleterious alleles (see below), it has not been enough to fully restore the loss of genetic diversity resulting from founder events occurred during the colonization process.

A second consequence of the northward range expansion in *L. longirostris* was an accumulation of genetic load (i.e., expansion load), as reflected in a higher proportion of putatively deleterious to non-deleterious mutations in *front* than *core* populations. This was true for both private and shared mutations, pointing to processes that took place during the range expansion itself (and excluding other demographic processes such as recent bottlenecks). The observed accumulation of deleterious relative

to non-deleterious mutations in the range front is consistent with the prediction of relaxed purifying selection in the expanding edge (e.g., Peischl et al., 2013). This is not supported, however, by the ratio of non-synonymous to synonymous mutations ($\theta_{\pi N}/\theta_{\pi S}$), i.e. the efficacy of selection, which was found to be similar at different parts of the species' range (Table 1.1). Interestingly, the discordance between these two quantities has been reported both in plants (*Mercurialis annua*, González-Martínez et al., 2017) and humans (reviewed in Lohmueller, 2014), having been attributed to the sensitivity of $\theta_{\pi N}/\theta_{\pi S}$ to non-equilibrium conditions (e.g., Gravel, 2016). In our case, such a discrepancy could also reflect the inclusion of splice sites located at the exon-intron boundaries in the calculation of genetic load. Although splice sites are crucial for proper splicing and some of them are under strong selective constraints, others are less conserved between species and evolve under weak selection, resulting in a substantial fraction of suboptimal nucleotides (genetic load) at these specific positions (Denisov et al., 2014).

Although significant, the levels of expansion load in *L. longirostris* were smaller than those found in other plant species. For instance, an increase of derived deleterious variants that are fixed in range-front populations have been reported for *Mercurialis annua* (González-Martínez et al., 2017) and *Arabis alpina* (Laenen et al., 2018). In contrast, the number of potentially deleterious mutations fixed in populations of *L. longirostris* was virtually non-existent, both in the *core* and the *front*. These differences are probably associated to the intensity of the bottlenecks experienced during the expansion, which were severe in the case of *M. annua* and *A. alpina*, and almost absent in *L. longirostris*.

There are several not exclusive factors that could have mitigated the severity of bottlenecks during range expansions in *L. longirostris*. First, the northward expansion has been a slow process, taking around 15,000 years, and occurred along a marked environmental gradient. The current climate difference between the temperate northern portion of the Iberian Peninsula and the drier central and southern parts has existed since the Middle Pleistocene (González-Sampériz et al., 2010), or even earlier (Jiménez-Moreno et al., 2010), so it is likely that individuals at the colonization front have been forced to adapt to novel conditions during the colonization process, slowing the rate of expansion. Recently, Gilbert et al. (2017) showed that when range expansions are slowed by the need to locally adapt, the severity of genetic drift is reduced, since slow

expansions provide more opportunities for high-fitness alleles to reach the colonization front through migration from the core, restoring genetic diversity and increasing the efficacy of selection. A slow colonization also facilitates gene flow among different sectors of the expansion axis; although we have shown that gene flow was not enough to restore genetic diversity in *front* populations, it may still have mitigated the impact of expansion load, preventing demographic collapse (Peischl & Excoffier, 2015; Peischl et al., 2016).

Second, experimental studies in both southern (Ruíz de Clavijo, 2001) and northern populations (García, 2004) of the species have shown that, despite its ability to self-fertilize, seed set is dramatically reduced after selfing, suggesting that strong self-incompatibility mechanisms have been maintained along the expansion axis. Thus, it is unlikely that populations in the colonization front were founded by a small number of migrants, as compatible crosses require distinct S-alleles (Allee effect). Allee effects have been found to increase effective population size at the front of expanding populations, lowering the role of genetic drift and gene surfing (Hallatschek & Nelson, 2008). In fact, theoretical models support that the maintenance of self-incompatibility in colonizing species is strongly linked to the evolution of high dispersal rates that compensate for their incapacity of founding new populations from single or few individuals (Pannell & Barrett, 1998; Dornier et al., 2008). Consequently, much of the genetic diversity initially lost by founder effects in colonization fronts could be restored, helping selection to lower the severity of expansion load. In this sense, a recent study in *Arabidopsis lyrata* ssp. *petraea* showed that, despite a sharp decline in effective population size during a postglacial colonization, the allelic diversity at the self-incompatibility locus has remained similar in *core* and *front* populations, suggesting that high migration rates have been promoted to avoid mate limitation (Takou et al., 2019a). Interestingly, the authors also found that the number of deleterious mutations in core and front populations has remained unchanged, suggesting that gene flow could also have buffered the effect of founder events on genetic load.

Third, *L. longirostris* is able to establish temporary seed banks, which can buffer genetic diversity losses. While the central achenes possess a well-developed pappus facilitating dispersal, peripheral ones exhibit low dispersal ability but have a thick pericarp that delay germination, so they can remain temporarily in the dormant state on the ground (Ruíz de Clavijo, 2001). Beyond optimizing reproductive success in highly

dynamic or disturbed habitats (Cohen, 1966), seeds stored in the soil could have contributed to minimize the loss of genetic diversity and the fixation of putatively deleterious mutations in *front* populations by increasing their effective population size and buffering some genotypes from local extinctions.

Expansion load and fitness

The impact of potentially deleterious mutations on individual performance and long-term persistence of range-front populations is difficult to predict. However, the fact that they have persisted for thousands of generations (i.e. > 20,000 yrs) suggests that they could have relatively small fitness costs. Alternatively, the accumulation of deleterious mutations observed in *front* populations could be an indirect effect of adaptation. Geographic clines in life-history traits commonly evolve when expansions occur across environmental gradients (e.g., Colautti & Barrett, 2013), and the steep environmental cline from South to North Iberia may have provided opportunities for adaptation during the species expansion. Paradoxically, strong positive selection may counteract purifying selection in neighboring genomic regions and thus maintain deleterious variants at higher frequency than expected from their detrimental fitness effect (Hartfield & Otto, 2011). The hitchhiking of deleterious alleles along with positively selected variants has been reported in a variety of organisms, including humans (Schrider & Kern, 2017), dogs (Marsden et al., 2016) and plants (Zhang et al., 2016), among others. The potential interaction between the accumulation of potentially deleterious mutations and local adaptation in *L. longirostris* deserves further exploration through common garden or reciprocal transplants, as well as the analysis of selective footprints at the molecular level.

Chapter 2

Environmental patterns of adaptation: phenology, size and biomass allocation

2.1 Introduction

Natural selection has commonly been considered to be the main force shaping genetic architecture in plant populations, usually leading towards local adaptation (Linhart & Grant, 1996). Thus, for instance, some important traits related to phenology, growth, size or cold tolerance have been shown to exhibit significant genetic variation shaped by selection along environmental gradients (see review in Rathcke & Lacey, 1985; Zhen & Ungerer, 2008; Montesinos-Navarro et al., 2012). More recent reviews, however, conclude that local adaptation is less common than initially thought, particularly in small populations, since stochastic processes may limit the efficacy of selection (Leimu & Fischer, 2008; Hereford, 2009).

Species and their populations are currently responding to the challenges imposed by global climate change through range shifts (Chen et al., 2011; McCarty et al., 2017). For this reason, the role of evolutionary processes leading to or constraining local adaptation after range expansion or shift, are receiving a renewed interest (Hoffmann & Sgrò, 2011; De Meester et al., 2018; Miller et al., 2020). From a theoretical point of view, different demographic and genetic processes might affect the ability of populations to adapt during range expansion, such as reduced genetic diversity in expanding populations (Eckert et al., 2008; Slatkin & Excoffier, 2012), extensive gene flow from core/ancestral populations (Kirkpatrick & Barton, 1997; Lenormand, 2002), or expansion load driven by gene surfing (Excoffier et al., 2009; Peischl et al., 2013). Furthermore, adapting to novel environmental conditions requires the integrated response in a complex phenotype, at least for some suite of functional and developmental characters (Murren, 2012; Laughlin & Messier, 2015). Therefore, antagonistic selective pressures among correlated traits in novel environments may further constraint adaptation (Etterson & Shaw, 2001; Blows & Hoffmann, 2005; Colautti et al., 2010).

Among the most important traits in sessile organisms like plants are those involved in phenology, the timing of events during the life cycle, since they determine the moment and duration of vegetative growth and reproduction (Wilczek et al., 2010). For instance, the timing of germination starts the vegetative growth phase and imposes the conditions under which seedlings will grow, and flowering initiates the reproductive phase where fertilization and maturation of fruits will occur. Phenological traits also condition other aspects of the life cycle, such as survival, size and/or reproductive

output, since they determine the circumstances in which resources will be acquired (Rathcke & Lacey, 1985; Donohue et al., 2010). Therefore, selection pressures on phenological traits are expected to be strong, particularly for plant species for which overall fitness depends on a single reproductive event.

Phenology itself might be considered a complex phenotype, and in natural populations selection is likely to occur on multiple components of phenology at the same time (Donohue, 2005; Galloway & Burgess, 2009; Galloway et al., 2018). These might include the season and timing of germination, and the time at and duration of flowering and fruiting. Phenological traits can be correlated in many ways (Peiman & Robinson, 2017). For instance, the timing of later life-history events can be contingent on environmental cues regulating earlier life stages (Donohue, 2002, 2005; Wilczek et al., 2010), therefore creating a functional linkage. Furthermore, there might also exist genetic correlations between different phenotypic traits sharing the same genetic pathways, e.g. the timing of germination and flowering (Chiang et al., 2009; Martínez-Berdeja et al., 2020). Plant phenological traits are usually studied in isolation, and some of them have received special attention from an evolutionary perspective. This is the case of flowering time, that is usually considered the most important phenotypic trait coordinating the life cycle with local environmental conditions, and therefore the main determinant of darwinian fitness in annual and perennial monocarpic plants (Metcalf et al., 2003; Roux et al., 2006). Selection factors affecting germination timing have also been extensively studied in many species (see, for instance, Rathcke & Lacey, 1985 and references therein), though the explicit consideration of its influence on the phenotypic expression and natural selection of other postgermination phenological or life-history traits have not been empirically addressed until quite recently (see Donohue et al., 2010). In fact, studies assessing the joint evolution of phenological traits in natural populations of non-model species are rather limited (but, see, for instance, Griffith & Watson, 2005; Prendeville et al., 2013; Kooyers et al., 2015; Galloway et al., 2018).

From a methodological point of view, reciprocal transplant experiments are the most powerful way to testing for local adaptation (Kawecki & Ebert, 2004). However, when dealing with a large number of populations, detection of clines along geographic and/or environmental gradients is an alternative valuable approach (Mayr, 1956; Endler, 1977; Montague et al., 2008; Prendeville et al., 2013, and references therein). Detecting geographical or environmental adaptive clines might depend on the nature, scale and

direction of selective pressures, as well as their interaction with the underlying genetic architecture (Galloway et al., 2018). For instance, in the model plant *Arabidopsis thaliana*, seed dormancy and flowering time co-vary negatively with minimum temperature (Debieu et al., 2013; Vidigal et al., 2016), but the strength of the correlation between the two traits is geographically variable, being especially constrained in southern latitudes (Marcer et al., 2017). Therefore, in order to gain relevant information on selective pressures shaping local adaptation, studies addressing clinal variation across the whole life-history should preferably be conducted on a wide geographic or environmental scale. Since non-adaptive evolutionary processes, mainly associated with demographic history, can also yield clinal patterns, a useful additional approach is testing among-population trait variability while controlling for selectively neutral processes (Keller et al., 2009; Kooyers et al., 2015). Finally, although every biological system or study in evolutionary ecology might be unique (Nakagawa & Parker, 2015), a powerful way to increase generality comes through comparisons between related or unrelated species with different evolutionary histories but sharing a common geographical and environmental setting.

In this study we examine the variation in phenological (season and time to germination and flowering) and other fitness-related traits (e.g., size at flowering, total biomass and biomass allocation) in two closely related Asteraceae species: *Leontodon longirostris* and *L. saxatilis*. Our study was conducted in common environmental conditions and included 42 populations of *L. longirostris* covering most of its natural range in the Iberian Peninsula, and six populations of *L. saxatilis*. In the Iberian Peninsula *L. longirostris* behaves mostly as an annual species and *L. saxatilis* as a polycarpic perennial. The geographical analysis of genetic variability and structure (de Pedro et al., 2021) strongly suggest that *L. longirostris*: 1) has undergone a south-north latitudinal range expansion starting around 40,000 years ago; 2) this expansion was accompanied by a loss of genetic diversity and an increase of genetic load along the expansion route; and 3) gene flow among populations is quite high. Our main objective is twofold: to test for local adaptation across populations of *L. longirostris* along its expanding range, and to compare patterns of phenotypic (co)variation between life-history traits among *L. longirostris* populations and between both species. We addressed the following particular questions: (i) is there any evidence for local adaptation in life-history traits in both taxa?; (ii) what is the influence of relevant phenological traits, such

as the timing of germination and flowering, on other life-history traits linked to fitness?; (iii) what is the relationship between variation in life-history traits and environmental factors?; and (iv) are our results consistent with those previously reported for the model plant *A. thaliana* in the same geographical setting? Our experimental approach includes testing for population differentiation along environmental clines after controlling for genetic structure arising from selectively neutral evolutionary processes.

2.2 Material and methods

Study species and field sampling

Leontodon longirostris (Finch & P.D. Sell) Talavera (≡ *Leontodon saxatilis* subsp. *rothii* Maire ≡ *Thrincia hispida* Roth) and *L. saxatilis* Lam. (≡ *Thrincia saxatilis* (Lam.) Holub & Moravec, ≡ *Leontodon taraxacoides* (Vill.) Mérat, nom. illeg.) are two closely related species of the Asteraceae family. Both are self-incompatible and outcrossing (Ruiz de Clavijo, 2001; García, 2004; personal observations based on experimental crosses). Plants grow forming a basal rosette and flower from March to August producing a variable number of flowering axes, each developing a single flower-head bearing several (20-180) ligulate flowers that are pollinated by generalist insects. After fecundation, two types of achenes are produced: "peripheral", without a well-developed pappus, and "central", with a well-developed pappus composed of 10-14 bristles (Ruiz de Clavijo, 2001; Brändel, 2007). Low germination percentages under experimental treatments, and differences between achene types, suggest some kind of primary dormancy, partly due to pericarp thickness (Ruiz de Clavijo, 2001). *Lentodon longirostris* is described as an annual or biennial plant that is naturally distributed in Western Europe (mainly in the Iberian Peninsula), north-western Africa and the Macaronesian region (Talavera et al., 2015). It grows in abandoned agricultural fields, annual grasslands, roadsides, and other disturbed habitats on a variety of soils. It can be found on a wide range of elevations, from sea level up to more than 1,500 m. *Leontodon saxatilis* is a widely distributed perennial lineage, occurring naturally in Eurasia and the Azores Islands (Talavera et al., 2015). It grows in more stable habitats such as alpine grasslands, peatlands and wetlands, and its distribution in the Iberian Peninsula is much more restricted to northern and Atlantic coastal areas.

During 2013 and 2014 we visited more than 100 localities of *L. longirostris* and 11 populations of *L. saxatilis* distributed across the Iberian Peninsula following a three-parallel south-north transect sampling scheme (Fig. 2.1). We sampled achenes from 42 localities of *L. longirostris* and six of *L. saxatilis* (Fig. 2.1, Table S2.1). Depending on the size of the population, we collected one flower-head from 15 to 50 individuals (maternal families) per population, located 2-5 meters apart, and we stored them at 4 ºC until their use in this study. During two of the sampling trips, by the end of May in 2013 and 2014, we also performed a visual quantitative estimate of the percentage of flowering plants in the populations visited (geographic locations are given in Table S2.1), according to the following qualitative numerical values: 1 (0-25%), 2 (25-50%), 3 (50-75%), and 4 (75-100%). Qualitative estimate values per population were based on average observations taken from two perpendicular transects.

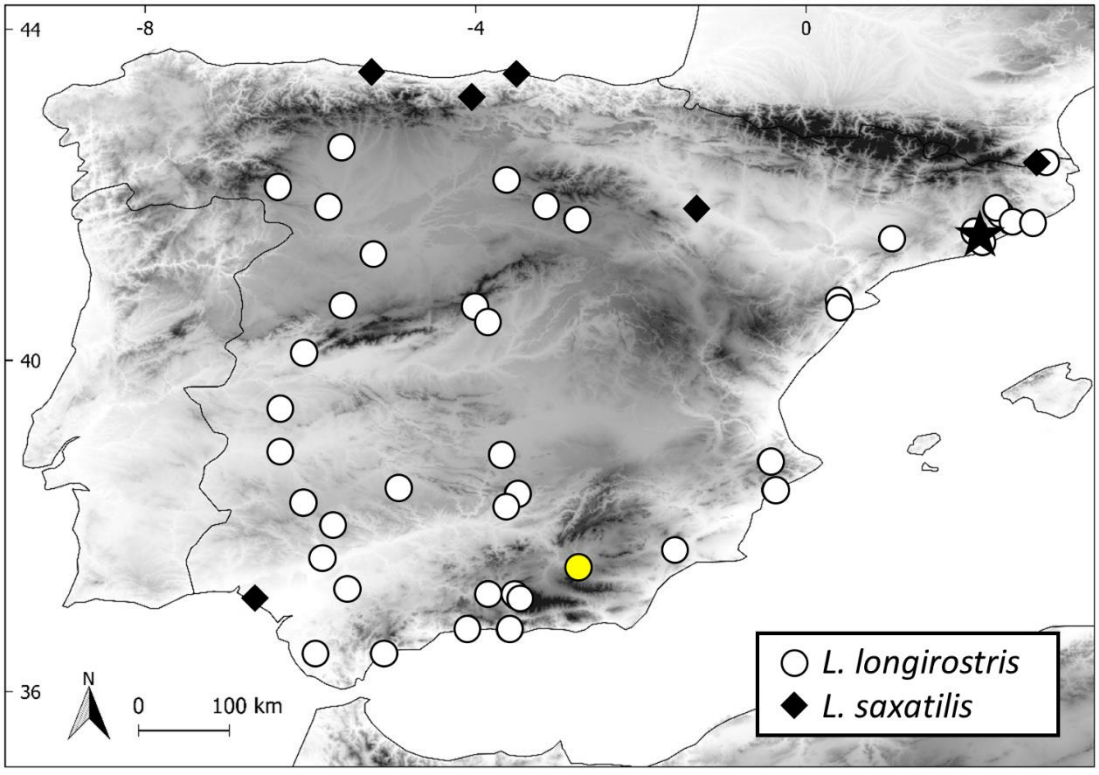


Figure 2.1. Location of *Leontodon longirostris* (circles) and *L. saxatilis* (diamonds) populations across the Iberian Peninsula. Population LL-51 is highlighted in yellow (see text for further explanation). Seeds collected at these sites were grown in a common garden located at the Autonomous University of Barcelona (blackstar). Elevation is indicated in the map from 0 m (white) to > 2,000 m (dark grey). Details of the populations are given in Table S2.1 (Supplementary Material).

Common garden experiments

Two common garden experiments were carried out in the experimental field station of the Autonomous University of Barcelona, north-eastern Spain (Fig. 2.1). In order to explore the maximum variation of the phenotypic traits considered across environmental conditions, we favored a high number of populations over the inclusion of many families per population.

A first experiment was performed by the end of September 2014. We sowed seeds from all the sampled *L. longirostris* populations (42) and five populations of *L saxatilis*. A total of 1,138 seeds from 15-25 families per population were sown in a randomized complete block design, each block containing one single representative per family and population. Populations with less than 25 families were not represented in all blocks. Since germination was rather limited for some of the populations, we only recorded the germination rate for two months. The same experiment was conducted by the end of winter the following year, in early March 2015. For this second experiment we tried to use the same populations and families as in the first one. This was not possible for two *L. longirostris* populations (LL-20 and LL-29), which were replaced by two other nearby populations sampled in the same location. We also included an additional *L. saxatilis* population (LS-11, see map in Fig. 2.1). In this second experiment we sowed a total of 1,168 seeds from 15-26 families using the same randomized block design.

In both experiments, achenes (seeds) were sowed individually into 6 x 6 x 5 cm pots filled with a 4:4:2 sand:silt:universal substrate (Batlle 960005UNID - Substrate Universal 80 L, Batlle, Molins de Rei, Barcelona) and placed in a greenhouse at room temperature. Only central achenes were used to avoid the differences in germination and seedling vigor that have been reported between central and peripheral achenes (Ruiz de Clavijo, 2001). Seeds emerging in the second experiment (591) were transplanted to 10 cm diameter x 10 cm height pots filled with the same substrate used for germination, and then transferred to a shade house keeping the structure of the randomized block design. Both seeds and plants were watered as needed using automatic sprinklers.

Besides assessing germination rate, in the second experiment we also measured several life-history traits covering the life cycle of the plants, from germination to reproduction and eventually, in many cases, to death. Three phenological traits, the time from sowing to seedling emergence (germination time), from seedling emergence

to flowering (flowering time), and from flowering to senescence (senescence time), were estimated by daily recording each individual achene and plant, respectively. Germination time was scored as the number of days from seeding to the emergence of cotyledons, flowering time as the number of days from germination to the emergence of the first flowering stalk, and senescence time as the number days from flowering to death or removal at the end of the experiment. The effect of phenological traits on plant performance was assessed estimating size at flowering and plant biomass. To calculate size at flowering we used a non-destructive method based on digital imaging. Each plant was photographed when the first flowering stalk appeared and size at flowering was estimated as the total rosette area using the open-source software ImageJ (Rasband, 1997-2018). Plants were harvested when they began to senesce or at the end of the experiment (November 2015). Upon harvesting, plant material was separated into leaves and flowering stalks to estimate above-ground biomass. The biomass of the flowerheads produced (including achenes, receptacle, and bracts) was estimated as the number of flowering stalks times the average flowerhead weight obtained from a subsample of 1-16 (median = 4) mature capitula. In addition, we estimated below-ground plant biomass in a subsample of 303 plants: roots were collected, separated from soil and weighted. Dry weight was estimated after plant material was oven-dried at 70 ºC for 48h. We also measured root-crown diameter (RCD) of all plants when harvested. RCD was the best predictor for total plant biomass: r^2 = 0.72 (n = 273) and 0.59 (n = 30), for *L. longirostris* and *L. saxatilis*, respectively, on log-transformed variables. Finally, we estimated the reproductive vs. vegetative biomass allocation (RVB: reproductive to vegetative biomass ratio) considering the estimated biomass of flowerheads produced as a measure of the resources allocated to reproduction, and the biomass allocated to leaves, flowering stalks and roots as a surrogate of vegetative growth.

Climate data

In order to assess the potential environmental selection agents behind population differences in life-history traits, we investigated the association between climatic variables and the phenotypic traits considered. For each population, 50-year average values of mean maximum and minimum daily temperature and precipitation were obtained from the Digital Climatic Atlas of the Iberian Peninsula (http://opengis.uab.es/wms/iberia/en_index.htm; Ninyerola et al., 2005). In addition to

mean values, we also considered the coefficient of variation in precipitation as an estimate of the among-year predictability in water availability. For subsequent analyses, climatic variables were grouped into two climatic seasons: spring (March, April and May), where most of growing and flowering is expected to occur, and summer (June, July and August), where most achenes are expected to develop under natural conditions.

Statistical analyses

We used linear mixed models to account for the different sources of variation using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2019). We explored both differences between species and among populations within each species. When testing for differences between species, populations were modelled as random effects and species as fixed effects. In models analyzing trait differences among populations within each species, populations were treated as fixed effects. All models included block identity as a random effect. Mean length and width of five achenes per maternal family were used as fixed covariates in all models to account for potential size-dependent variability in initial resources. We conducted analyses on germination rate assuming a binomial error distribution, whereas for the rest of traits we considered a gaussian error distribution after conveniently transforming (log or square root) the original data. The significance of fixed effects in models with binomial errors was obtained from likelihood ratio tests, while for models with gaussian errors we used F-tests with Satterthwaite's approximation for the degrees of freedom implemented in the *lmerTest* package (Kuznetsova et al., 2017). Since neither achene length nor width had a significant effect in any of the models, they were removed from the final analyses. Population least squares means for each trait were obtained after adjusting the models using the *emmeans* package (Lenth, 2019). Depending on the trait considered, analyses were repeated after excluding one population (*L. longirostris*: LL-51) which showed clear extreme values (see Results).

To determine the potential role of environmental conditions in explaining trait differences among *L. longirostris* populations we used Pearson's correlations between adjusted population trait means and climatic variables. To assess the potential confounding role of neutral demographic processes arising during or after geographic expansion (e.g., drift, migration and founder events) on adaptive phenotypic clines, we

used genetic markers as a null model. We estimated genetic similarity from the first two axes of a Principal Component Analysis using 168,733 SNPs from a targeted DNA sequencing analysis (de Pedro et al., 2021). Following Keller et al. (2009), to assess the potential effect of neutral processes generating the observed phenotypic trait clines, we used a model selection approach based on the Akaike information criteria (AIC) among three candidate linear models. For each individual trait we compared a "Climate" model which included only climatic explanatory variables, an "Ancestry" model using only the first two PCA axes, and a "Climate + Ancestry" model that included both types of explanatory variables. This comparison was performed using 20 *L. longirostris* populations for which both phenotypic and genetic data were available (only 17 for RVB).

Finally, the phenotypic covariation among the adjusted population trait means (germination rate, germination time, flowering and senescence times, size at flowering, RVD, and RCD) was quantified with Pearson's correlation using all sampled populations of *L. longirostris*, excluding the outlier population LL-51. Correlations involving RVB were explored in a subset of 35 populations for which this information was available. We also used individual (log or square root transformed) data to perform a path analysis to better summarize the relationships between life-history traits and to measure their relative direct and indirect contribution on plant size (RCD). The model was constructed for each species separately to show the sequential effect of phenology (germination and flowering times) on size at flowering, senescence time, and RCD. Path analyses were performed using the *lavaan* package (Rosseel, 2012).

2.3 Results

Trait differences between species and populations

We found significant differences for all the traits considered between the two taxa (see "Taxa *P*-value" and mean estimates in Table 2.1). Values of germination rates were much lower for *L. longirostris* than *L. saxatilis* in both germination experiments, particularly during the early fall experiment (Fig. S2.1A and B, Table 2.1). While mean germination rate for *L. saxatilis* was quite similar in both experimental conditions (89.7% and 92.0% during early fall and late winter, respectively), mean germination rate for *L. longirostris* populations was much lower during fall (8.1%) than during late

winter experimental conditions (44.3%). In the late winter experiment, plants of *L. longirostris* germinated later (two days on average; Fig. S2.1C, Table 2.1) and flowered earlier (15 days on average; Fig. S2.1D, Table 2.1) than those of *L. saxatilis*. When reaching sexual maturity, the size (rosette area) of *L. saxatilis* plants doubled that of *L. longirostris* (Fig. S2.1E, Table 2.1), also attained larger RCD values at the end of the growing period (Fig. S2.1F, Table 2.1). As expected, some individuals and populations of the short-lived *L. longirostris* behaved as annuals, becoming senescent as early as 1-5 months after flowering (Fig. S2.1G, Tables 2.1 and S2.2), while most *L. saxatilis* individuals behaved as perennials. During the period considered, allocation to vegetative and reproductive growth also varied significantly between the two lineages: *L. saxatilis* invested on average about half of the resources in reproduction and less than twice when compared to *L. longirostris* (Fig. S2.1H, Table 2.1). Inspection of population averages for some of the traits in the *L. longirostris* populations sampled, particularly germination rate, flowering time, size at flowering, RCD, and RVB, showed that one of the populations (LL-51) behaved as a clear outlier (Fig. S2.1, Table S2.2). Similar results (not shown) were obtained when excluding this population showing extreme values from the between-species comparisons.

There were also significant differences for almost all traits within both taxa (see "Pop *P*-value" in Table 2.1). Values for all traits and populations are given in Table S2.2. As mentioned above, one population of *L. longirostris* (LL-51: Baza), showed much higher germination rates, flowered much later (71 days on average), reached a larger size at flowering and RCD at harvesting, and invested less in reproduction than the other *L. longirostris* populations, resembling in fact the perennial *L. saxatilis* plants (Fig. S2.1, Table S2.2). When excluding this outlier population from the analyses, differences among *L. longirostris* populations for all traits except germination rate in the early fall experiment remained significant (Table 2.1). We also found significant differences among *L. saxatilis* populations for all the traits, except those related to germination rate, though differences during the first (early fall) experiment were found to be marginal (Table 2.1).

Table 2.1. Common garden estimates of marginal means for trait values ± SE in *Leontodon longirostris* and *L. saxatilis*, and statistical significance for comparisons between species (Taxa *P*-value) and populations within species (Pop *P*-value) inferred from linear mixed models. The total number of populations (N-Pop) and families (N-Fam) are indicated for each trait. *P*-values obtained for the analyses excluding population LL-51 are given in brackets. RCD: root crown diameter; RVB: reproductive to vegetative biomass ratio.

		Leontodon longirostris				***Leontodon saxatilis***			
	Taxa *P*-value	**N-Pop**	**N-Fam**	**Mean ± SE**	**Pop *P*-value**	**N-Pop**	**N-Fam**	**Mean ± SE**	**Pop *P*-value**
Germination rate 1- Early fall experiment (%)	**< 0.001**	42	1,016	8.1 ± 0.01	**< 0.001** (0.189)	5	122	89.7 ± 0.01	0.074
Germination rate 2- Late winter experiment (%)	**< 0.001**	42	1,020	44.3 ± 3.23	**< 0.001 (<0.001)**	6	148	92.0 ± 3.01	0.583
Germination time (days)	**< 0.05**	42	457	14.0 ± 0.30	**< 0.001 (<0.001)**	6	134	12.0 ± 0.70	**< 0.01**
Flowering time (days)	**< 0.05**	42	435	62.0 ± 2.30	**< 0.001 (<0.001)**	6	119	77.0 ± 70	**< 0.001**
Rosette size at flowering (cm^2)	**< 0.01**	42	434	16.0 ± 2.00	**< 0.001 (<0.001)**	6	117	40.0 ± 8.0	**< 0.001**
RCD (mm)	**< 0.001**	42	435	2.5 ± 0.10	**< 0.001 (<0.001)**	6	119	3.7 ± 0.3	**< 0.001**
Senescence time (days)	**< 0.01**	42	435	140.0 ± 3.70	**< 0.001 (<0.001)**	6	118	171.0 ± 8.40	**< 0.001**
RVB	**< 0.001**	36	273	1.2 ± 0.07	**< 0.001 (<0.01)**	6	30	0.5 ± 0.12	**< 0.001**

Prior to the analysis, variables were ln- (germination and flowering time) and square-root transformed (rosette size at flowering, RCD, senescence time and RVB).

Relationships between traits

Results of pair-wise trait correlations based on average population values for *L. longirostris* are shown in Table 2.2. Germination rates estimated in the two experiments were not significantly correlated to any of the other traits measured ($P > 0.05$), though germination rate during early fall showed a positive and marginally significant ($r = 0.3$, $P = 0.066$) association with RCD, and germination rate in late winter was positively and marginally significantly correlated with the size at flowering ($r = 0.27$, $P = 0.089$). Late germinating seeds produced plants that were smaller than those germinating earlier. Late flowering phenotypes were significantly bigger than early flowering ones (Table 2.2). However, we did not find a significant correlation between germination and flowering times, although phenotypes that took longer to germinate tended to flower earlier. We also found a strong significant positive correlation between size at flowering and RCD ($r = 0.88$, $P < 0.001$; Table 2.2). On the other hand, both phenological traits (germination time and flowering time) had a significant effect on RCD, i.e. final plant size and biomass (Table 2.2). Senescence time was not correlated with any of other trait. Plants flowering earlier allocated more biomass to reproduction than late-flowering ones (Table 2.2). Similar results were obtained when using Spearman's correlation coefficients (data not shown). Pair-wise correlations between traits in *L. saxatilis* were not estimated because of their limited sample size.

Table 2.2. Pearson's correlation coefficients (*r*) between mean trait values (estimated marginal means from linear mixed models) in *Leontodon longirostris* populations grown in common environmental conditions. Population LL-51, with extreme values, was excluded from the analyses. Germination rates 1 and 2 correspond to early fall and late winter experiments, respectively. RCD: root crown diameter; RVB: reproductive to vegetative biomass ratio. *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; † $P < 0.10$; *ns = not significant*.

	Germination rate 1	**Germination rate 2**	**Germination time**	**Flowering time**	**Rosette size at flowering**	**RCD**	**Senescence Time**
Germination rate 2	0.26 *ns*						
Germination time	-0.20 *ns*	-0.23 *ns*					
Flowering time	0.06 *ns*	0.18 *ns*	-0.21 *ns*				
Rosette size at flowering	0.15 *ns*	0.27 †	-0.65 ***	0.75 ***			
RCD	0.30 †	0.24 *ns*	-0.53 ***	0.72 ***	0.88 ***		
Senescence time	0.25 *ns*	-0.02 *ns*	0.22 *ns*	0.09 *ns*	-0.08 *ns*	0.17 *ns*	
RVB[a]	0.02 *ns*	-0.24 *ns*	-0.18 *ns*	-0.44 **	-0.26 *ns*	-0.29 †	-0.18 *ns*

[a]data available only for 35 pops.

Based on individual plant data, the path diagrams in Fig. 2.2 summarize de sequential direct and indirect relationships among some of the traits considered, from germination time to the final size achieved (RCD), and they also provide a useful comparison between both species. Size at flowering was the trait with the strongest effect on RCD (standardized path coefficients: 0.770 and 0.718, for *L. longirostris* and *L. saxatilis*, respectively). The variability in senescence time also showed a significant direct effect on the total biomass achieved in both taxa, though the effect was larger in the case of *L. saxatilis* (Fig. 2.2). Flowering time in *L. longirostris* plants showed an important positive indirect effect (0.278) on RCD through size at flowering. *L. saxatilis* also showed the same indirect effect, albeit much larger (0.449). In the case of *L. longirostris*, there was a small direct negative effect (-0.096) of germination time on the total size achieved (RCD), but also a strong indirect negative effect through size at flowering (-0.513) and, therefore, a total negative effect on RCD (-0.609). In the case of *L. saxatilis*, germination time also showed a negative, albeit smaller, indirect effect on RCD through size at flowering (-0.221), but also a positive indirect effect through flowering time and size at flowering (Germination time → Flowering time → Size at Flowering → RCD: 0.209), and a negative indirect effect through flowering time and senescence time (Germination time → Flowering time → Senescence time → RCD: -0.034), so that its total effect on RCD was negligible (-0.046). The model for *L. longirostris* accounted for 71% of the total observed variance in RCD, whereas that for *L. saxatilis* accounted for 60% of the variance.

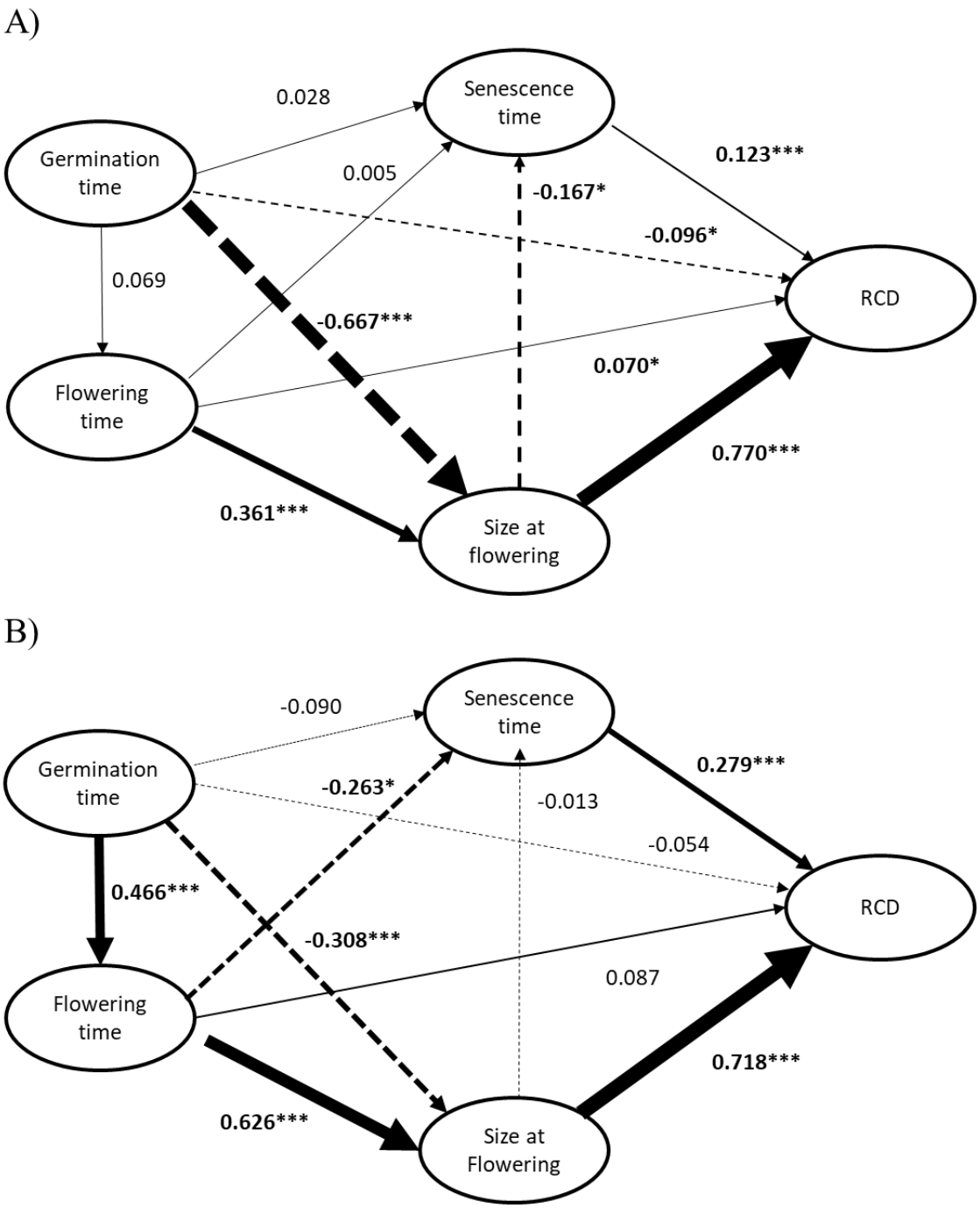


Figure 2.2. Path diagrams showing the direct and indirect sequential effects of traits on RCD (root crown diameter) for each taxon: **A**) *Leontodon longirostris*; **B**) *L. saxatilis*. Line thickness stands for magnitude of path coefficients. Population LL-51, with extreme values, was excluded from the analyses. Solid and dashed lines indicate positive and negative effects, respectively. Significant path coefficients: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.

Trait-climate associations

In *L. longirostris*, among-population variability in most of the examined traits was significantly correlated with some of the environmental variables considered (Table 2.3). Although germination rates were not associated with any of the climatic variables, absolute differences in germination between both experiments were significantly and negatively related to summer precipitation (Fig. 2.3). This negative trend was reinforced when pooling data from both species. Germination time was mainly related to spring and summer minimum temperatures: seeds from warmer sites tended to germinate later than those from colder ones (Fig. 2.4A and B, Table 2.3). Flowering time in the common environment was positively correlated with spring and summer precipitation, and negatively with minimum and maximum temperatures and among-year precipitation variability, both during spring and summer (Fig. 2.4C and D, Table 2.3), whereas senescence time was only positively associated with summer precipitation. Flowering phenology (flowering scores) estimated in the field also showed significant and pronounced climatic clines, with populations from warmer and drier environments showing more advanced phenological states than those from colder and wetter sites (Fig. S2.2). Size at flowering and RCD showed significant correlations with temperature and precipitation: plants from warmer and drier sites flowered earlier and were smaller than those from colder and wetter populations (Table 2.3). RVB showed significant positive and negative correlations with temperature and precipitation, respectively: plants from hot and dry habitats investing more in reproduction. In general, though, for most traits, spring and summer temperatures showed higher correlations than precipitation-related variables, highlighting the role of temperature behind the observed geographical variability.

Model comparisons of climatic variables with ancestry-derived measures of genetic structure suggested that selectively neutral demographic processes were not the main factor underlying the observed phenotypic clines with climate (Table S2.3): models with only climatic predictive variables showed the lowest AIC values, except for senescence time, for which the lowest AIC was that of the Climate + Ancestry model. Finally, though we did not test for associations between mean population trait values and climatic variables in *L. saxatilis* because of the reduced sample size, some traits, in particular germination time, also appeared to vary according to a latitudinal trend, following precipitation and temperature gradients (see Fig. S2.1).

Table 2.3. Pearson correlation coefficients (*r*) between mean trait values (estimated marginal means from linear mixed models) and environmental variables for *Leontodon longirostris* populations grown in common environmental conditions. Population LL-51, with extreme values, was excluded from the analyses. Germination rates 1 and 2 correspond to early fall and late winter experiments, respectively. RCD: root crown diameter; RVB: reproductive to vegetative biomass ratio. *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; † $P < 0.10$; *ns = not significant*.

	Germination rate 1	**Germination rate 2**	**Germination time**	**Flowering time**	**Rosette size at flowering**	**RCD**	**Senescence time**	**RVB**
Spring min. temp.	-0.24 *ns*	0.02 *ns*	0.47 **	-0.72 ***	-0.68 ***	-0.71 ***	0.06*ns*	0.06 *ns*
Summer min. temp.	-0.19 *ns*	-0.05 *ns*	0.49 **	-0.74 ***	-0.74 ***	-0.72 ***	0.14*ns*	0.14 *ns*
Spring max. temp.	-0.15 *ns*	0.06 *ns*	0.33 *	-0.68 ***	-0.70 ***	-0.71 ***	0.09*ns*	0.22 *ns*
Summer max. temp.	-0.05 *ns*	0.06 *ns*	0.05 *ns*	-0.54 ***	-0.55 ***	-0.57 ***	-0.04*ns*	0.45 **
Spring prec.	0.06 *ns*	0.03 *ns*	-0.08 *ns*	0.54 ***	0.44 **	0.54 ***	0.10*ns*	-0.35 *
Summer prec.	0.20 *ns*	-0.06 *ns*	-0.03 *ns*	0.66 ***	0.47 **	0.65 ***	0.36 *	-0.41 *
Spring prec. variability (CV)	-0.10 *ns*	0.11 *ns*	0.36 *	-0.65 ***	-0.55 ***	-0.56 ***	0.10*ns*	-0.04 *ns*
Summer prec. variability (CV)	-0.21 *ns*	0.16 *ns*	0.13 *ns*	-0.64 ***	-0.39 *	-0.54 ***	-0.18*ns*	0.08 *ns*

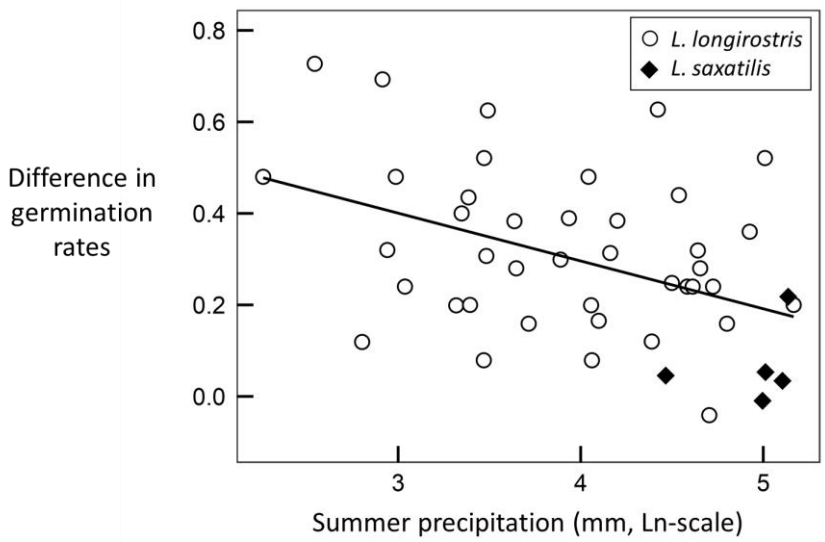


Figure 2.3. Relationship between the difference in germination rates (late winter - early fall experiments) and summer precipitation in *Leontodon longirostris* (circles) and *L. saxatilis* (diamonds). Pearson's correlations: r = -0.31 (P < 0.05) and r = -0.43 (P < 0.01) for *L. longirostris* and pooled data for both species, respectively.

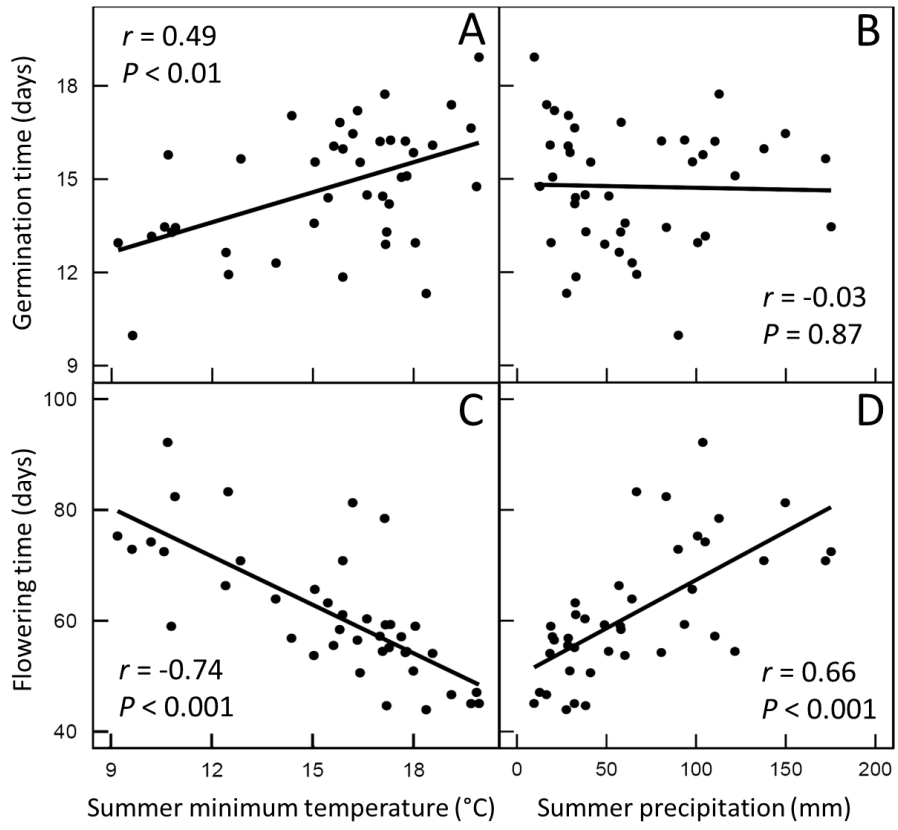


Figure 2.4. Relationships between phenological traits (germination and flowering times) and summer environmental variables in *Leontodon longirostris*. **A**-**B**) Germination time; **C**-**D**) Flowering time. Points are population expected marginal means from linear mixed models obtained in the common garden. Pearson's correlation coefficients (r) and associated P-values are indicated in each plot.

2.4 Discussion

In this study we evaluated the extent of variation in life-history traits in two non-model species (*L. longirostris* and *L. saxatilis*) and, in particular, the effect of the timing of phenological events on fitness-related traits. Our results indicated substantial population differentiation for most of the traits analyzed, both between and within species. Overall, we found a continuum of life-history trait variation in the timing of germination and duration of the life cycle. In general, *L. longirostris* showed more restricted conditions for germination and shorter vegetative and reproductive phases, germinating later and flowering earlier than *L. saxatilis*, which resulted in an ample reduction of the total size and higher investment in reproductive structures. In the case of *L. longirostris*, most traits were associated with climatic clines of temperature and seasonal drought, suggesting that this variation was the result of adaptation to specific environments. Although maternal effects cannot be completely ruled out (Mousseau & Fox, 1998; Wolf & Wade, 2009), our results support theoretical models and observations indicating that shorter life cycles evolve as an adaptive response to more unpredictable and stressful environments, in agreement with other empirical data (reviewed in Friedman, 2020). Furthermore, model comparisons of trait-climate associations including neutral genetic structure suggest that other potential confounding factors, such as past demographical processes, do not fully contribute to explaining the patterns observed.

Both theoretical and empirical evidence give support for the seasonal timing of germination as one of the most crucial developmental phases influencing natural selection on other life-history traits, particularly in short lived and annual species (Donohue, 2002; Burghardt et al., 2015). We found strong differences in germination behavior between species. Seeds from *L. saxatilis* showed high germination rates under different experimental conditions (i.e., late fall and early winter), while those of *L. longirostris* showed lower germination rates, especially under early fall conditions. This result suggests that *L. longirostris* seeds have developed differential sensitivity to environmental cues for germination or breaking dormancy. Range-wide variability in the timing of germination linked to primary and secondary dormancy has also been found in *Arabidopsis thaliana* (e.g., Debieu et al., 2013; Martínez-Berdeja et al., 2020). Whereas in *A. thaliana* this variability is significantly associated with latitude or climate and interpreted as a fine-tuned adaptation to climate and/or habitat disturbances, in *L. longirostris* we did not find any significant relationship between germination rate and

climate. However, we did find a significant relationship between the difference in germination rate between experimental conditions and summer rainfall. This association between variability in germination and water availability suggests a strategy for sensing optimal conditions for germination and reduce fitness variance in unpredictable environments (i.e., diversified bet-hedging, Venable, 2007). A high environmental (thermal) dormancy induction associated with a gradient of increasing temperatures and decreasing precipitation is found in *A. thaliana* populations along an altitudinal cline in northeastern Spain (Montesinos-Navarro et al., 2012) and across Europe (Kronholm et al., 2012), a mechanism that can help to avoid early germination and seedling death after sporadic rainfall events, very frequent in the Mediterranean climate.

Flowering time had a strong direct effect on final plant size in both species, whereas germination time (i.e., the time for seedlings to emerge) had also a strong indirect effect in the case of *L. longirostris*. Population differentiation for both phenological traits was evident in both taxa. Within *L. saxatilis*, variation was mainly due to the only population located in the south (LS-11), that clearly exhibited earlier flowering, smaller size, and a greater investment in reproduction than populations located in northern latitudes. In fact, this population showed similar values for both traits compared to many *L. longirostris* populations. Although *L. saxatilis* is mainly a perennial lineage, populations from southwestern Iberia can behave as annuals and/or biannuals (Talavera et al., 2015), in accordance with our results. The shift between annual and perennial strategies is a frequent, although still not well understood process, and seems to involve cis-acting changes in Flowering Locus C (*FLC*) orthologues that are differentially expressed (Kiefer et al., 2017). Thus, the particular environmental conditions occurring in the south of the Iberian Peninsula might have favored the expression of an annual life cycle in a perennial lineage.

In the case of *L. longirostris*, we found substantial differentiation and a clear latitudinal cline for phenological traits, giving rise to a continuous range of variation in life cycles, from short-lived annual plants to long-lived perennial ones. Delayed germination, advanced flowering, and partly also early senescence, characterized short-lived plants from de south, inhabiting warmer and drier habitats. Similar latitudinal patterns were observed in the field for flowering phenology, which was clearly advanced in southern populations at the time of sampling. In general, organisms are expected to evolve towards a reduction in lifespan to maximize their fitness in stressful

and/or spatially and temporally unpredictable environments (Stearns et al., 2000; Griffith & Watson, 2005; Cotto & Ronce, 2014; Friedman, 2020). Experimental studies simulating natural selection under water stress in the model species *A. thaliana* during four generations have shown a fast selection for shorter life cycles (Brachi et al., 2012). Similarly, a 5-year drought affecting populations of the annual plant *Brassica rapa* in southern California caused a shift to earlier flowering and at a smaller size, supporting that phenological traits can be selected very fast (Franks & Weiss, 2008). In *L. longirostris*, the higher temperatures and limited water availability characterizing southern populations in the Iberian Peninsula compared to those from the north, strongly reduce the length of the favorable growing season, likely imposing strong selection for drought escaping strategies favoring early flowering to ensure that reproduction is achieved before summer drought. Similar range-wide patterns of variation as those found in *L. longirostris*, with early flowering and smaller phenotypes located in environments with short favorable growing seasons, have been often reported and attributed to local adaptation in other annual plants (e.g., Debieu et al., 2013; Brouillette et al., 2014; Schneider & Mazer, 2016). Seedlings of *L. longirostris* phenotypes with shorter lifespan from warmer and dryer habitats also tended to emerge later (larger germination times). Early germination strategies might be particularly favored in more stable and competitive environments (e.g., Ross & Harper, 1972), but not in variable and unpredicted habitats. In fact, the survival benefit of early emergence seems to be greater in perennials than in annuals, since perennials in predictable habitats can bet that all their offspring will emerge in the best conditions (Verdú & Traveset, 2005).

The phenological patterns described above jointly determine overall plant size and development and, potentially, also fitness. They also give rise to a range of plant sizes following an environmental cline. Delayed emergence of seedlings, early vegetative-reproductive stage transition, and advanced senescence of phenotypes as a consequence of a shorter favorable growing season comes at the cost of smaller size, as well as lower vegetative and reproductive biomass. Nevertheless, comparisons between species showed that the short-lived *L. longirostris* allocated roughly twice of its biomass to reproduction compared to the perennial *L. saxatilis*. Similarly, within *L. longirostris*, differences in biomass allocation were also related to annuality: short-lived plants from warmer and drier habitats allocated more biomass to reproduction than to vegetative

growth, an expected strategy to maximize fitness in plants reproducing once in a lifetime. Higher reproductive effort is commonly found in annual plants from arid, unpredictable environments as compared to conspecific plants growing under more productive conditions, that should invest more into survival (e.g., Aronson et al., 1993; Petrů et al., 2006; Kurze et al., 2017).

Contrary to other studies (e.g., Vidigal et al., 2016; Martínez-Berdeja et al., 2020), we do not find a significant relationship between the timing of germination (either germination rate or germination time) and flowering time. Germination and flowering are expected to be intimately related: seed germination often responds to the environmental cues (e.g., temperature, photoperiod) experienced by maternal plants during flowering and seed maturation through different mechanisms that modulate the intensity of seed dormancy (Donohue, 2009; Imaizumi et al., 2017). Conversely, seeds appear to be able to integrate abiotic cues that influence adult phenology (e.g., flowering) in response to the environment experienced during the embryo stage, regardless of their dormancy state (Rubio de Casas et al., 2012). This is because both germination and flowering share important genetic and metabolic pathways, causing environmentally induced pleiotropic effects on both developmental phases. In *A. thaliana*, for example, the flowering gene *FLC* has been found to regulate seed germination (Chiang et al., 2009), while the dormancy gene *DOG1* also affects flowering time (Huo et al., 2016; Martínez-Berdeja et al., 2020). Furthermore, the flowering time locus *FLC* might also pleiotropically control additional morphological and physiological traits, some of them related to water use efficiency (see review in Kalisz & Kramer, 2008). In our study, the lack of significant relationship between germination and flowering might be due to limited population sampling and poor within-population replication. An alternative explanation is that different genetic co-variation patterns exists and/or are expressed between both traits depending on the population considered and the particular environmental condition used for experimentation. Variable covariation patterns, following a latitudinal trend, between seed dormancy, vegetative growth and flowering times, have been reported in the model species *A. thaliana* at the European scale (Debieu et al., 2013). Similarly, in the same species, flowering time and seed dormancy co-vary negatively with minimum temperature in the Iberian Peninsula: accessions from warmer locations exhibit higher seed dormancy and flower earlier. However, the covariation pattern among both traits

differs among regions, being particularly constrained in the southwestern area where minimum temperatures are highest (Marcer et al., 2017). These results suggest that the temperature-driven relationship between seed dormancy and flowering time might be variable, and that environmental boundaries exist for the co-evolution of both traits. In any case, additional information regarding environmental cues affecting germination and dormancy are needed in order to explore similar co-variation responses in *Leontodon*.

Of particular interest is the comparison of adaptation patterns in the major life-history transitions between unrelated taxa sharing a similar habitat and the same history of environmental change and selective pressures. The comparison is even more interesting when the taxa do not share the same demographic history. While the overall history of *L. longirostris* in the Iberia Peninsula is dominated by a progressive expansion of the species from south to north (de Pedro et al., 2021), the presence of several divergent lineages in *A. thaliana* in the same region suggests a complex population dynamics, consistent with the existence of multiple Pleistocene refugia (Picó et al., 2008). Interestingly, in spite of these different demographic histories, patterns of selection on the main life-history traits between both taxa quite mirror each other. In both species, flowering time is a relevant trait influencing the whole life cycle, and natural variation in this trait seems to be largely mediated by temperature (Méndez-Vigo et al., 2013; Manzano-Piedras et al., 2014; Marcer et al., 2017), giving rise to faster life cycles in southern locations. Furthermore, the variability found in the germination and dormancy patterns of the seeds in *A. thaliana* (Vidigal et al., 2016), and that found for the variability in the germination rates in *L. longirostris*, were both associated with locations exposed to little summer rainfall. The fact that two phylogenetically unrelated species distributed throughout the same geographical area responded in a similar way to the same environmental drivers highlights the relevant role of adaptive processes for persistence and in determining species' range sizes.

Finally, the evidence reported here on local adaptation throughout most of the *L. longirostris'* range sheds some additional light into the relationship between demography and adaptive processes. de Pedro et al. (2021) showed that the south-north range expansion across de Iberian Peninsula resulted in a loss of genetic diversity and the accumulation of potentially deleterious mutations (expansion load) in range-front populations. Theoretical work suggests that expansion load could prevent adaptation to

novel environments and limit a species' range (Peischl et al., 2015). The levels of expansion load found in *L. longirostris* were rather low compared to other plant species, probably due to a combination of factors, including high levels gene flow among populations and a self-incompatible breeding system. In addition, since the expansion occurred along a strong environmental cline, it has been suggested that adaptation to novel environments could have played a key role to slow down the expansion, preventing the fixation of deleterious variants in the range-front (de Pedro et al., 2021). Our results point to this direction, because the observed levels of expansion load do not seem to have prevented local adaptation to the environmental variation encountered across its range. Processes actually constraining the current geographical range of the species could probably be associated to other evolutionary constrains, such as limited genetic diversity and/or genetic correlations among traits (see, for instance, Kalisz & Kramer, 2008).

Taken together, our results suggest that local adaptation exists for phenological traits in response to environmental varying selection, and that southern *L. longirostris* populations have evolved to face short growing periods by selecting traits favoring short life cycles. In addition, we found evidence supporting that variation on phenological traits has a relevant role in explaining the variability in other traits (e.g., size) closely linked to fitness. However, further experimental studies, using a combination of ecological and molecular approaches, are needed to fully understand the mechanisms underlying selective processes leading to adaptation in this species.

Chapter 3

Unraveling the genetic basis for adaptation during range expansion

3.1 Introduction

Range expansions are common in nature, and they involve numerous eco-evolutionary processes (Miller et al., 2020). Identifying and understanding the interactions between demographic, ecological and genetic processes facilitating or limiting such events is a central topic in evolutionary biology. At the genetic level, range expansions often result in a progressive loss of genetic diversity along the expansion route due to founder events (Austerlitz et al., 1997). Also, because selection is less efficient, deleterious alleles may accumulate and reach high frequencies at the range-edge, reducing the fitness of individuals ("expansion load", Peischl et al., 2013; Peischl & Excoffier, 2015). Both the loss of genetic diversity and expansion load can interact in a complex way with the adaptive dynamics of expanding populations (e.g., Gilbert et al., 2017), which could slow down expansions and limit species' ranges (Peischl et al., 2015). Because many range expansions take place along environmental gradients, understanding these interactions is then essential to explain why some species fail to expand while others easily spread across different habitats.

The limitation of small populations for evolutionary change has been extensively studied, both from the theoretical and the empirical point of view (reviewed in Willi et al., 2006). Regarding expansion load, though there is increasing evidence for the accumulation of deleterious alleles along expansion routes in several organisms (e.g., Lohmueller et al., 2008; Henn et al., 2016 in humans; Grossen et al., 2020 in Alpine ibex; González-Martínez et al., 2017; Laenen et al., 2018; Willi et al., 2018; de Pedro et al., 2021 in plants), few studies have examined to date the detrimental effect that the excess of deleterious alleles can have on the adaptive potential of populations. In plants, two studies have shown a link between an increased mutation burden and reduced adaptation at range margins. In the North American plant *Arabidopsis lyrata* L., postglacial expansions were accompanied by a strong increase in mutation load at the range-edges, which in turn resulted in reduced rates of population growth and reproduction (Willi et al., 2018). Similarly, in recently expanded western European populations of *Mercurialis annua* L., the accumulation of mildly to strongly deleterious mutations was linked to a significant reduction in the number of selective sweeps compared to ancestral populations from Greece and Turkey (González-Martínez et al., 2017).

By contrast, other studies in plants suggest that rapid evolution to novel environmental conditions has occurred when the species expanded their ranges in response to current climate change (e.g., Colautti & Barrett, 2013; Lustenhouwer et al., 2018). Recent theoretical studies predict a reduction of expansion load when range expansions occur along environmental gradients (Gilbert et al., 2017). In these conditions, the expansion rate at the leading edge is expected to slow down by the need for colonizing populations to adapt to the new environments, allowing more time for migrants to arrive and for selection to reduce the frequency of deleterious mutations (Gilbert et al., 2017). On the other hand, changing selection pressures in the expanding fronts can promote the spread of pre-existing beneficial variants that rapidly sweep to high frequencies in the population, counteracting the negative effect of expansion load, as has been proved experimentally in microbial communities (Bosshard et al., 2019; Gralka et al., 2016).

Standing genetic variation (SGV) provides the potential for rapid adaptation to novel conditions (Barrett & Schluter, 2008; Messer & Petrov, 2013). Standing variation allows beneficial alleles to be readily available for selection at the time that conditions change, so there is no need to wait for new advantageous mutations to arise (Barrett & Schluter, 2008). Then, given that the success of expansions is directly linked to the ability to adapt to novel environments at short timescales, we could expect a predominant role of standing variation on adaptation in expanding populations. In this situation, adaptation should commonly produce "soft" selective sweeps, where multiple copies of the adaptive alleles sweep through the population at the same time (Hermisson & Pennings, 2005; Messer & Petrov, 2013). New beneficial mutations are less likely to drive rapid adaptation in the populations at the range-edge, because the small population size of founding populations limits their probability to appear. Nevertheless, if new beneficial variants arise, the exponential growth rates in the expanding front could provide good opportunities for spread and fixation (Otto & Whitlock, 1997), resulting in "hard" sweeps.

Leontodon longirostris (Finch & P.D. Sell) Talavera (≡ *Leontodon saxatilis* subsp. *rothii* Maire ≡ *Thrincia hispida* Roth) (Asteraceae, Cichorieae) is a common weed in the Western Mediterranean Basin. The demographic history of the species in the Iberian Peninsula supports a range expansion from the south starting around 40,000 yrs BP, and reaching the north around 25,000 yrs BP (de Pedro et al., 2021). The expansion was

accompanied by a remarkable loss of genetic diversity and a significant increase in the proportion of deleterious to non-deleterious mutations (expansion load) at the range-edge (de Pedro et al., 2021). However, the levels of expansion load were smaller than those found in other plant species, and none of the deleterious variants present in the expansion front was fixed at the population level (de Pedro et al., 2021). Among the factors that may explain these results are the high dispersal capacity and the self-incompatible mating system of the species, which could have mitigated the severity of bottlenecks during the expansion (de Pedro et al., 2021).

Besides, since the expansion of *L. longirostris* in the Iberian Peninsula occurred along a gradient of decreasing temperature and increasing precipitation towards the north, adaptation to novel environments during the expansion could have played a key role to prevent the fixation of deleterious variants in the range-front. Common garden experiments of *L. longirostris* populations distributed across the Iberian Peninsula have shown substantial genetic variability for phenological and other fitness-related traits, giving rise to a continuous range of variation in life cycles, from short-lived annual plants to long-lived perennial ones (de Pedro et al., unpublished). In particular, individuals from northern populations germinated earlier and flowered later that those from the south, suggesting that longer life-cycles have evolved in response to more favorable environments during the expansion (de Pedro et al., unpublished). The change in life-cycle was associated with a greater plant size and biomass, as well as lower investment in reproduction, suggesting that adaptation to local climate was mediated by the co-variation among functionally related traits (de Pedro et al., unpublished). However, since the experiments were performed under a unique environmental setting from field-collected seeds, plastic responses and maternal effects cannot be completely ruled out (Kawecki & Ebert, 2004).

Integrating information from both genomic and phenotypic approaches can provide stronger evidence for local adaptation, while allowing to identify genes underlying specific phenotypes (Savolainen et al., 2013; Villemereuil et al., 2016). Combined with detailed information on climatic variables, such an approximation can also reveal relevant associations between putative adaptive loci and environmental drivers of selection. In this study, we used several genomic approaches to complement the phenotypic information obtained in the common garden experiments to better understand the potential adaptive processes that occurred during the northward

expansion of *L. longirostris* in the Iberian Peninsula. We have taken advantage of the recently available targeted sequencing design used to infer the demographic history and expansion load of the species, consisting of ~1.59 Mbp sequenced in 238 *L. longirostris* plants from 21 populations (de Pedro et al., 2021). This genomic dataset included almost 300 genes, some of them well-known candidates for adaptation in other Asteraceae of commercial interest. We aimed to: (i) investigate the interaction between genetic diversity, expansion load and adaptation in populations of the expanding front; (ii) ascertain the relative contribution of standing and new genetic variation (e.g., soft vs. hard selection) to adaptation; (iii) identify candidate genes underlying phenotypic variation that could be the target of selection in the novel environments, encountered during the expansion; and (iv) detect and/or confirm climatic drivers relevant for selection and adaptation in *L. longirostris*.

3.2 Material and methods

Sampling, DNA extraction, and SNP genotypes

We used the same genotypic dataset designed to infer the demographic history, the genetic diversity and expansion load of *L. longirostris* in the Iberian Peninsula (de Pedro et al., 2021). Leaves from 6 to 20 plants (mean = 11) of the species were collected from 21 localities (Fig. 3.1, Table S3.1), resulting in a total sample size of 238 individuals. All localities but one (population 83) were included in the two common garden experiments carried out to obtain phenotypic data (see details in de Pedro et al., unpublished, and below). Seventy-eight of the 238 individuals collected for genotypic data were obtained by sowing seeds from the same maternal families used in both experimental settings (Table S3.1). In addition to the *L. longirostris* samples, 20 individuals of the close relative *L. saxatilis* Lam. (≡ *Thrincia saxatilis* (Lam.) Holub & Moravec, ≡ *Leontodon taraxacoides* (Vill.) Mérat, nom. illeg.) were collected from two localities in northern Iberia and used as an outgroup when needed. High quality genomic DNA was isolated from 50 to 100 mg of dry leaf material using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following standard protocols.

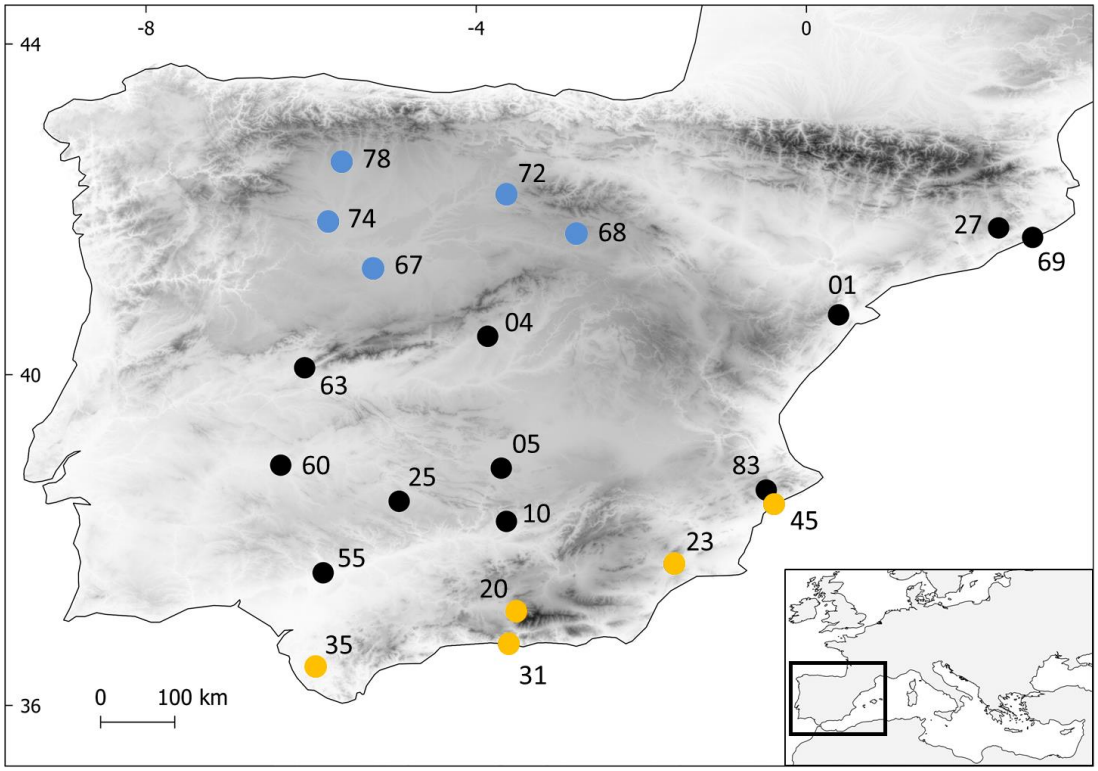


Figure 3.1. Geographic distribution of the 21 *Leontodon longirostris* populations included in this study. The two groups of populations used to identify selective sweeps in the *core* and the *front* areas are indicated by orange and blue dots, respectively. Details on the location, sample size, phenotypic and climatic features for each population are provided in Table S3.1.

SNP genotypes were obtained by targeted sequencing as described in de Pedro et al. (2021). Briefly, we resequenced 143 scaffolds (~1.59 Mbp) containing 299 annotated genes, most of them (259) with a functional ortholog in related Asteraceae species (e.g., *Lactuca sativa* L., *Helianthus annuus* L., *Artemisia annua* L., *Cynara cardunculus* L.). Only the larger 92, with lengths ≥ 10 kbp, contained complete genes. Approximately one third of the scaffold sequences corresponded to genic (coding and non-coding) regions (0.58 Mbp), while the remaining two thirds were intergenic (upstream and downstream) zones (1.01 Mbp). After sequencing, SNPs were called using GATK, following best practices, and further filtered using VCFTools v0.1.15 (Danecek et al., 2011) and vcffilter (vcflib C++ library), allowing for a maximum level of missing data per SNP of 10% and a maximum level of missing information per individual of 25% (see details in de Pedro et al., 2021). Only biallelic SNPs were retained, resulting in a final dataset of 168,733 SNPs distributed along intergenic regions (63%), exons (13%) and non-coding sections of genes (24%). Each SNP was annotated using SNPEFF v4.3t

(Cingolani et al., 2012) according to the predicted effect of the variant change based on the *L. longirostris* genome draft as a reference.

Phenotypic trait variability

Phenotypic data for 20 of the 21 genotyped populations were retrieved from the two common garden experiments performed in de Pedro et al. (unpublished). Both experiments included 42 *L. longirostris* and 6 *L. saxatilis* populations distributed across the Iberian Peninsula. The first experiment was carried out at the end of September 2014, and was used to estimate the proportion of seeds germinating in early fall for each population (germination rate 1). The experiment included a total of 1,138 seeds from 15-25 families per population that were sown in a randomized complete block design, each block containing one single representative per family and population (populations with less than 25 families were not represented in all blocks). The second experiment was settled by the end of winter of the following year, in March 2015, and was used to assess the germination rate in late winter (germination rate 2), as well as several life-history traits covering the life cycle of the plants: germination time (from sowing to seedling emergence), time and size at flowering, root-crown diameter (as a predictor for total plant biomass), and reproductive to vegetative biomass ratio. This second experiment included a total of 1,168 seeds from 15-26 families and used the same randomized block design than the first experiment. After germination, 591 seedlings were transplanted keeping the structure of the randomized block design (see details in de Pedro et al., unpublished).

Detection of selective sweeps in core- and front-range populations

Two complementary approaches were used to identify regions with signatures of selective sweeps. On the one hand, the composite-likelihood-ratio (CLR) test implemented in SweeD v3.3.4 (Pavlidis et al., 2013) was applied to detect strong positive signals of selection (i.e., 'hard' selective sweeps) from the site frequency spectrum (SFS). On the other hand, we computed the G12 and G2/G1 statistics developed by Harris et al. (2018) in order to detect hard and soft sweeps from multi-locus genotypes. In both analyses, the calculations were performed on two groups of populations, a first group including populations from the core distribution of the species

(south of the Iberian Peninsula), hereafter *core*, and a second one including populations from the colonization front, *front* (Fig. 3.1, Table 3.1). We used groups instead of single populations to have a greater sample size, which improves the accuracy of selective sweep detection (Pavlidis et al., 2013). Five populations in each group were selected according to their geographical position and similar genetic composition (i.e., the five populations in each group constitute a single gene pool, see Fig S1.1), and individuals were pooled together. The *core* group was composed by the five southernmost populations (20, 23, 31, 35, 45), with a total size of 68 individuals, while the *front* group was formed by the closest five populations of the expansion front (67, 68, 72, 74, 78), with a total size of 57 individuals (Fig. 3.1). The *core* group was characterized by a significantly higher genetic diversity and lower genetic load than the group of populations of the expansion *front* (Table 3.1). In addition, both groups represented two extremes within the south-to-north climatic gradient characterizing the Iberian Peninsula. Populations from lower latitudes (*core*) were warmer and drier than those from the *front* (minimum temperature (ºC) = 6.54 ± 3.23 vs. -0.57 ± 0.81; maximum spring temperature (ºC) = 20.39 ± 1.67 vs. 15.84 ± 1.22; coefficient of variation of spring precipitation = 0.81 ± 0.10 vs. 0.56 ± 0.06; summer precipitation (mm) = 23.28 ± 9.17 vs. 83.16 ± 21.54, respectively). Selection analyses were run separately for each group and based on individual scaffolds, using only those containing complete genes (i.e., 92 scaffolds with length ≥ 10 kbp, 233 annotated genes), which involved 104,362 and 81,902 SNPs in the *core* and *front*, respectively.

Table 3.1. Average genetic diversity and genetic load (± standard deviation) for the two groups of populations used for detection of selective sweeps (based on 116,946 SNPs with known allelic state; retrieved from de Pedro et al., 2021). *N*: sample size; Het: percentage of heterozygous sites; θ_π: nucleotide diversity (Tajima, 1983); Additive genetic load: proportion of deleterious mutations to non-deleterious mutations; Recessive genetic load: proportion of deleterious mutations to non-deleterious mutations in homozygous state.

Group	***N***	**Het**	**θ_π (x10^3)**	**Additive genetic load**	**Recessive genetic load**
Core	68	4.08 ± 0.32	3.32 ± 0.18	0.59 ± 0.03	0.41 ± 0.01
Front	57	3.00 ± 0.43	2.33 ± 0.24	0.64 ± 0.05	0.57 ± 0.14

CLR tests were computed using SWEED with default parameters except for the grid size that was fixed to 15-bp (http://pop-gen.eu/wordpress/software/sweed). Alternative grids with 30-bp and 50-bp produced similar results. Monomorphic sites were included in the analyses to improve the accuracy of the algorithm in estimating the location of selective sweeps (Pavlidis et al., 2010). It also makes the tests more robust to demographic processes, such as bottlenecks (Crisci et al., 2013). Ancestral vs. derived states were inferred by comparison with *L. saxatilis* and both folded and unfolded sites were considered in the analyses. Significance thresholds for selective sweeps were obtained by running SWEED on simulated data sets under specific demographic scenarios (without selection) for each population group. We inferred the best-fitting demographic parameters for *core* and *front* with FASTSIMCOAL2 v2.6.0.3 (Excoffier et al., 2013) using a simple three-epoch model (Fig. 3.2). Mutation rate, prior parameter distributions and run lengths were obtained from de Pedro et al. (2021). Once the optimal models were identified, MLCOALSIM v1.42 (Ramos-Onsins & Mitchell-Olds, 2007) was used to simulate 100 sequences per scaffold under each specific model. Simulated sequences were obtained without recombination and considering the same number of segregating sites (SNPs) than the observed data. A CLR test was considered significant if the observed value was greater than any of those obtained in the simulations.

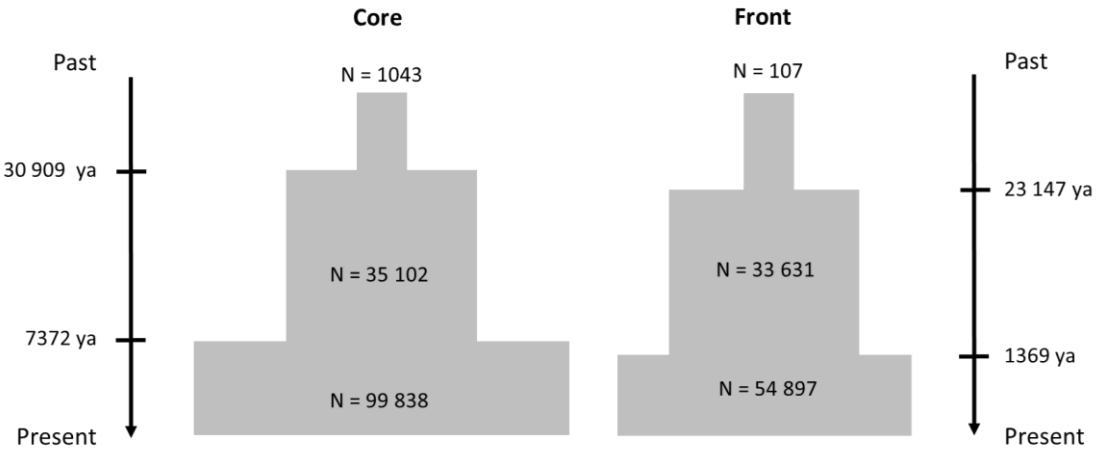


Figure 3.2. Schematic representation and parameter values of the two optimal demographic models inferred with FASTSIMCOAL2 for the *core* and *front* populations. Both models were used to obtain neutral sequences with coalescent simulations performed with MLCOALSIM. See text for further details.

G12 and G2/G1 statistics were calculated using the public python scripts available at https://github.com/ngarud/SelectionHapStats. G12 is the expected multi-locus genotype (MLG) homozygosity in which the frequencies of the first and the second most common MLGs are combined into a single frequency (Harris et al., 2018). Pooling the two largest MLG frequencies would provide no additional power to detect hard sweeps, but because soft sweeps rise at least two (and often more) MLG to high frequencies, distortion of their joint distribution allows for a stronger signal for soft sweeps (Harris et al., 2018). Compared with previous methods based on haplotype homozygosity, G12 does not require prior phasing of the genotype data, typically unavailable for non-model organisms. To gain insight about the "softness" of the sweeps detected, we calculated the G2/G1 statistic, where G1 is the MLG homozygosity and G2 is the MLG homozygosity calculated using all but the most frequent MLG. When sweeps are soft, the frequencies of the first- and second-most frequent MLG are both large, and the exclusion of the most frequent MLG should not decrease G2 value to the same extent than under hard sweeps, resulting in higher G2/G1 ratios. G12 and G2/G1 statistics were calculated using a window size of 51 SNPs and a step size of 10 SNPs, corresponding to averaged sequence lengths of 667 bp and 841 bp for *core* and *front*, respectively. This window size is large enough to minimize local elevated MLG homozygosity not resulting from a selective sweep, and small enough to avoid diluting selection signals due to mutation and/or recombination. In other outcrossing Asteraceae, such as wild sunflower (*Helianthus annuus* L.), linkage disequilibrium has been reported to decay to negligible levels at distances below 300 bp (Liu & Burke, 2006). To assess whether observed G12 values were elevated relative to neutral expectations, we run G12 on the same set of simulated sequences used to obtain significant thresholds in SWEED. Because G12 take unphased diploid MLGs as input, and MLCOALSIM outputs haploid individuals, we manually merged two random pairs of haploid sequences into a single individual. Thresholds for significant tests were defined by the highest G12 obtained for the 92 simulated scaffolds under each demographic model (threshold for *core*: 0.232, threshold for *front*: 0.205). Contiguous sequence windows with significant tests were considered to represent single selection events.

Genotype-phenotype (GPA) and genotype-environment (GEA) association

In addition to the tests described above that rely solely on genetic data, we searched for significant associations between the whole SNP dataset (168,733) and phenotypic traits, as well as with specific climatic variables. Apart from complementing the results obtained with methods designed to identify selective sweeps, with this approach we intended to detect and/or confirm climatic drivers relevant to adaptation in *L. longirostris*, as well as to identify candidate genes underlying the phenotypic variation observed in the common garden experiments (see de Pedro et al., unpublished, and above).

To investigate the association between genetic and phenotypic data, we selected four potentially adaptive traits explaining most of the variability observed for the species in common environmental conditions (de Pedro et al., unpublished). Germination rates (germination rate 1, germination rate 2) were chosen because they defined marked differences among populations regarding their germination behavior: some populations were able to germinate during early fall and late winter, while others only germinated under the late winter environment, suggesting different strategies for sensing the optimal conditions for germination (de Pedro et al., unpublished). We also selected two phenological traits (germination and flowering times) because their variation determined a range of life cycles: plants from warmer and drier environments were characterized by shorter lifespans (determined by late germination and early flowering) than those from colder sites, suggesting distinct adaptive strategies linked to the predictability of the environment (de Pedro et al., unpublished). For each trait, averaged values per population were retrieved from de Pedro et al. (unpublished), except for population 83 that was not included in the experiment (Table S3.1). We also obtained individual values for the 78 maternal families included in the genotyping design.

Regarding climatic variables, for each population we obtained a 50-year monthly average values of maximum and minimum daily temperatures and precipitation from the Digital Climatic Atlas of the Iberian Peninsula (http://opengis.uab.es/wms/iberia/en_index.htm; Ninyerola et al., 2005). In addition to mean values, we also calculated the coefficient of variation in precipitation as an estimate of the among-year predictability in water availability. For subsequent analyses, climatic variables were grouped into four climatic seasons: winter

(December-February), spring (March-May), summer (June-August), and autumn (September-November). Eight climatic variables were used in association tests: minimum temperature of winter, minimum and maximum temperature of spring, precipitation of winter, spring, summer and autumn, and the coefficient of variation of spring precipitation (Table S3.1).

We used two association methods that take into account population genetic structure to minimize false positive rates and that allow for both GPA and GEA: BAYESCENV v1.1 (Villemereuil & Gaggiotti, 2015) and BAYPASS v2.1 (Gautier, 2015). BAYESCENV is based on the F_{ST} model but, contrary to other existing approaches, incorporate locus-specific genetic and environmental effects to take into account other processes different from local adaptation (Villemereuil & Gaggiotti, 2015). BAYPASS explicitly account for the shared history of populations under study by computing a covariance matrix across population allele frequencies (Gautier, 2015). BAYESCENV was run for each of the standardized variables with default parameters (20 pilot runs with 5,000 iterations and 5,000 MCMC samples after a burn-in of 50,000 iterations with 10 steps between each sample). In BAYPASS, standardized variables were run using default parameters (20 pilot runs with 1,000 iterations and 1,000 MCMC samples after a burn-in of 5,000 iterations with 25 steps between each sample) under the standard model (STD in Gautier, 2015). After running both methods, we retained all significant SNP associations with a climatic variable or a phenotypic trait with a *Q*-value < 0.1 in BAYESCENV and a Bayes Factor (BF) > 5 in BAYPASS. Within this dataset, we identified as best candidate genes those with significant SNPs located on coding regions. In addition, to account for SNPs in regulatory regions, we also considered significant associations in nearby gene-flanking (2 kbp upstream and downstream) and intronic regions. In order to evaluate the relative contribution of standing and new genetic variation to adaptation, outlier SNPs were classified as standing variants when both alleles were present in the *core* and *front*, and as *de novo* mutations when they were private from one of the two groups.

Annotation of candidate genes

Genes overlapping significant selective sweeps and/or with outlier SNPs were queried against the NCBI non-redundant protein sequences database with BLASTX, using a

minimum threshold of 40% of sequence identity and E-value < 1E-10. Then, we only retained annotations with a functionally characterized homolog in related model species with available full-genome reference sequences within the Asteraceae (e.g., *Lactuca sativa* L., *Helianthus annuus* L., *Artemisia annua* L., *Cynara cardunculus* L.).

3.3 Results

Selective sweeps in core- and front-range populations

We identified a total of 151 selective sweeps based on CLR and G12 tests (Fig. 3.3, and Tables S3.2 and S3.3). Most of them (128) were exclusive to one of the two regional groups. Of these, 49 were detected in the *core*, whereas the remaining 79 were exclusive from the *front* populations (Fig. 3.3). The number of exclusive sweeps detected by CLR and G12 strongly differed from one group to the other. While the CLR tests conducted with SWEED detected twice as many sweeps in the *core* (33) as in the *front* (17), the opposite was true for G12, where the number of significant sweeps in the *front* was more than threefold (66) that in the *core* (19). This suggests different modes of selection (soft sweeps vs. hard sweeps) being predominant in the two regional groups (see below). In addition, both methods also differed in the number of selective sweeps shared between *core* and *front* (3 vs. 18 for CLR and G12, respectively).

Only a few number of sweeps were encountered by both methods (Fig. 3.3). Three sweeps detected by SWEED in the *core* were also identified by G12 in the *front* or where shared between the *core* and *front*, suggesting that some sweeps remained undetected in one of the two groups due to a weak signal. Four exclusive sweeps found using G12 in the *front* were confirmed by SWEED (Fig. 3.3). Two of them were among the 15 top selective sweeps (Table S3.3) and involved two genes putatively encoding the proteins *enhanced disease resistance 2* (EDR2) and *1-aminocyclopropane-1-carboxylate synthase 7* (ACS7).

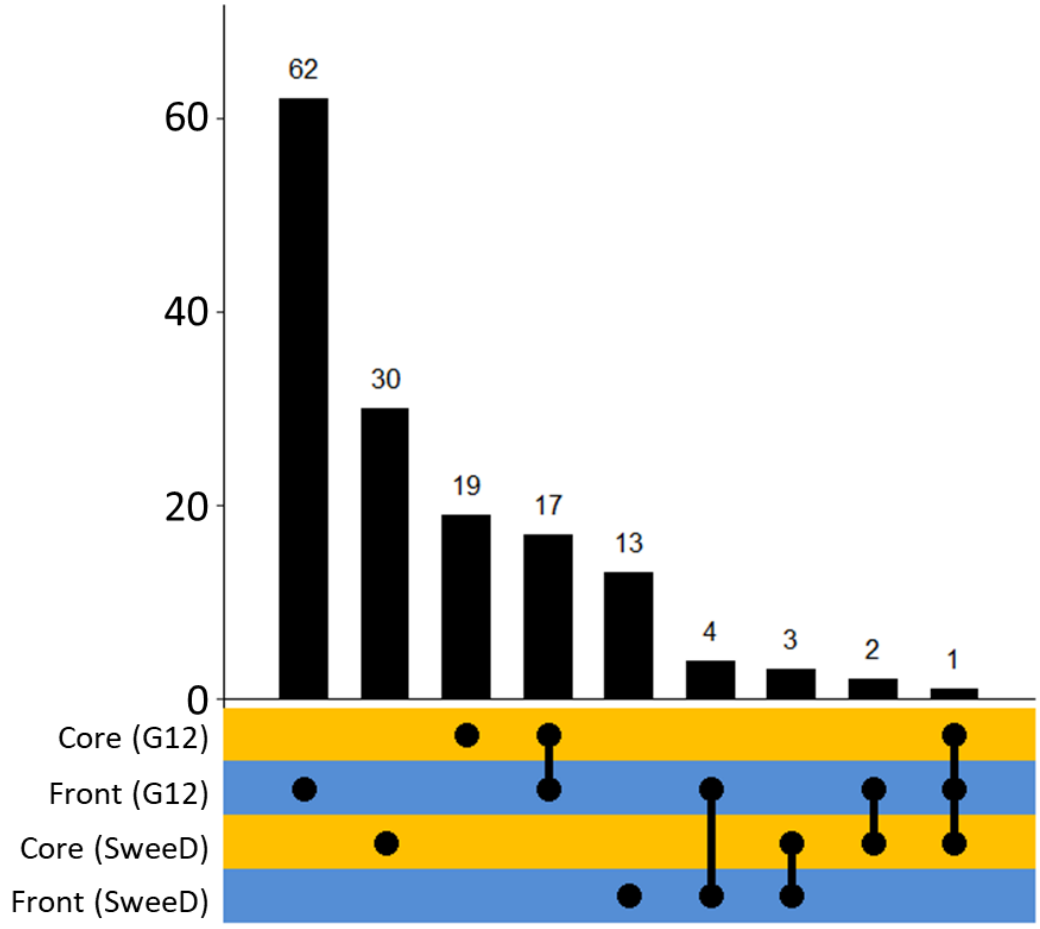


Figure 3.3. Number of significant selective sweeps obtained in the *core* and *front* populations using either SWEED or G12. Dark circles indicate the method (SWEED or G12) and geographical location (*core* or *front*) for significant sweeps, whereas vertical bars show the number of shared sweeps between different methods and/or geographical areas.

The MLG spectra of the top 15 selective sweeps exclusive from *core* and *front* populations revealed the presence of multiple genotypes in both datasets (Fig. 3.4). Some sweeps were characterized by two or more frequent MLGs, consistent with the signature of soft selective sweeps, while others had a single dominant MLG, consistent with the signature of hard sweeps. The presence of distinct MLGs at high frequency was not due to the mixing of distinct populations in a single group, since they were generally found in most of the five populations merged in the *core* or *front* groups (Fig. S3.1). There was a higher prevalence of soft sweeps in the *front* and of hard sweeps in the *core*: only one of the top 15 exclusive selective sweeps in the *front* had G2/G1 values below 0.1, consistent with harder sweeps, while seven had values above 0.25, in accordance with softer sweeps (Fig. 3.4). In contrast, the opposite pattern was found in the *core*, with eight sweeps showing G2/G1 values below 0.1 and only one above 0.25. This trend was maintained when all exclusive selective sweeps were taken into account (G2/G1 *core*: 0.127 ± 0.071, G2/G1 *front*: 0.256 ± 0.141; ANOVA, $P < 0.001$; Fig. S3.2), and also in those shared among both groups, although marginally significant (G2/G1 *core*: 0.133 ± 0.081, G2/G1 *front*: 0.205 ± 0.136; ANOVA, $P = 0.062$; Fig. S3.3). This points to different levels of "softness" depending on distribution range for selective sweeps affecting the same genome regions.

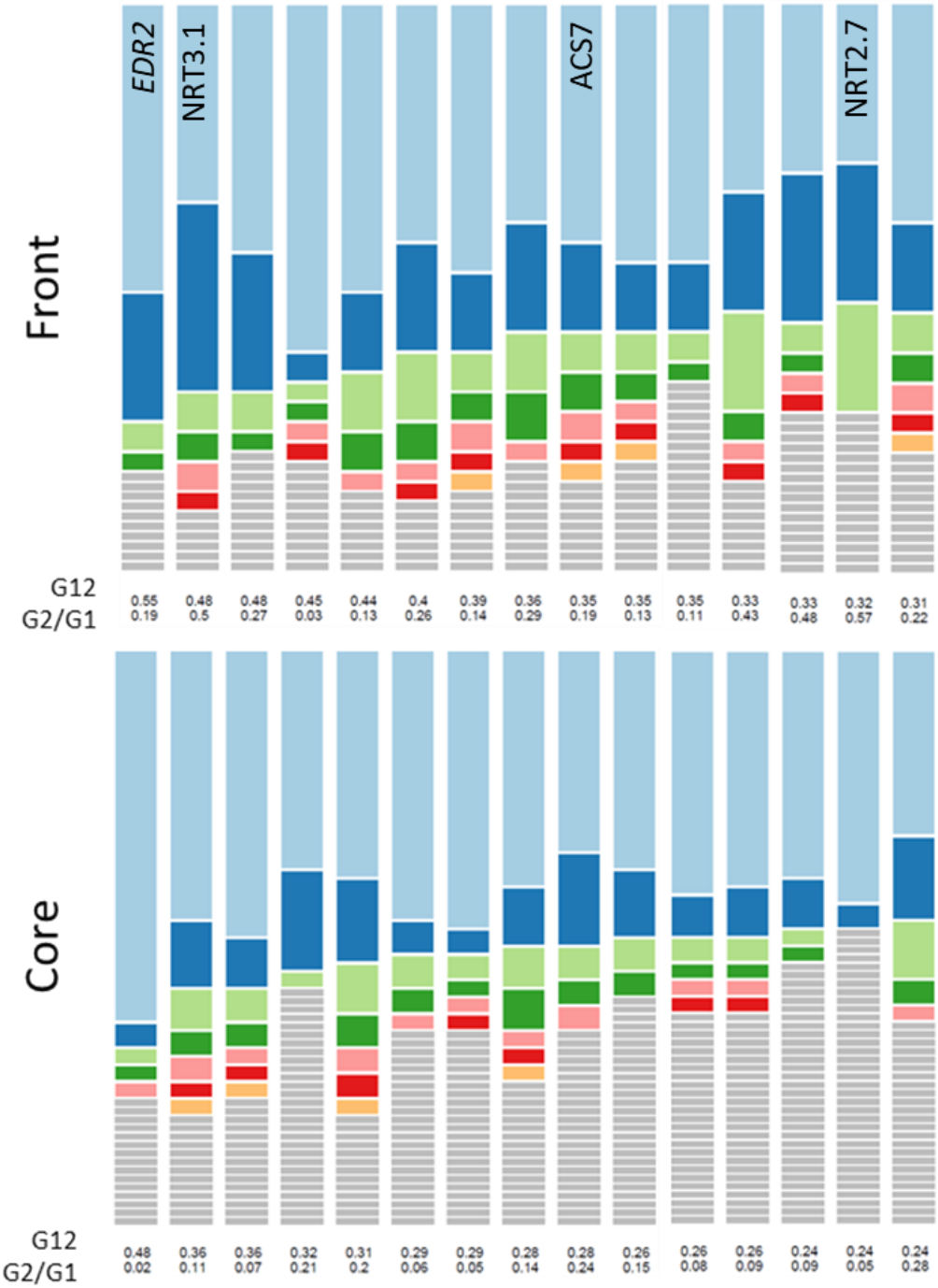


Figure 3.4. Multi-locus genotype (MLG) frequency spectra for the top 15 exclusive sweeps detected in *core* and *front* with G12 method. The height of the top light blue region in each bar indicates the frequency of the most prevalent MLG in the sample (68 and 57 individuals in *core* and *front*, respectively). Heights of subsequent colored bars indicate the frequency of the second, third, and so on, most-frequent MLG in a sample. Gray bars indicate unique MLGs. G12 is the expected MLG homozygosity combining the frequencies of the first and the second most common MLGs, and G2/G1 is the ratio between expected MLG including (G1) and excluding (G2) the most frequent MLG (see text for further details). Genes with significant G12 sweeps and outlier SNPs are indicated in regular type, while those with significant G12 and SWEED sweeps are shown in italics.

SNP association with phenotypic traits and climate

A total of 53 distinct SNP outliers, involving seven candidate genes and its surrounding regions, were found in common by BAYESCENV and BAYPASS (Table 3.2). The number of SNP outliers per gene and flanking regions ranged from 1 to 15, and included stop-gain, non-synonymous, synonymous, intron, 5'/3' UTR and intergenic variants. Most of the SNPs (46) were associated with a single variable, although a few showed significant correlation with two (5) or even three (2) of them. In all SNP loci showing signals of selection, both SNP variants were present in the *core* and *front*, likely representing standing variation.

Genotype-phenotype associations, GPAs

No significant associations were found for the germination rate 2 (late winter experiment) and the germination time. The germination rate obtained during the early fall experiment (germination rate 1) was significantly associated with a single synonymous variant located on a putative gene encoding for *5-methylthioadenosine/S-adenosylhomocysteine nucleosidase 2* (MTN2) (Table 3.2).

Flowering time was significantly associated with three SNPs, all located on the same gene, a putative *1-aminocyclopropane-1-carboxylate synthase 7* (ACS7) (Table 3.2). Two of the outliers were synonymous variants located on the first exon (SNPs S1 and S2, at positions 2,041 bp and 2,185 bp, respectively), while the third was located at the upstream region of the gene, at position 3,391 bp, very close to the starting codon (Fig. 3.5). A large significant selective sweep spanning from 2,503 to 9,388 bp was also detected for this gene in *front* populations with the G12 method (Fig. 3.4, Table S3.3), and affected the remaining four exons and an upstream region of about 6,000 bp (Fig. 3.5). A significant selective sweep involving part of this upstream region (around 9,000 bp) was also evidenced by CLR tests in SWEED, again only in *front* populations (Fig. 3.5, Table S3.2).

Table 3.2. Candidate genes with significant outlier SNPs detected with BAYESCENV and BAYPASS in *Leontodon longirostris*. ACS7: *1-aminocyclopropane-1-carboxylate synthase 7-like* (XP_023746525.1); NRT3.1: *High-affinity nitrate transporter 3.1-like* (XP_023758600.1); PHYE: *Phytochrome E-like* (XP_023758124.1); DRP5A: *Dynamin-related protein 5A-like* (XP_023761723.1); NRT2.7: *High affinity nitrate transporter 2.7* (XP_023728700.1); ML5: *Protein MEI2-like 5* (XP_023758128.1); MTN2: *5-methylthioadenosine/S-adenosylhomocysteine nucleosidase 2-like* (XP_023738120.1). Codes for SNP variants: T, stop-gained; N, non-synonymous; S, synonymous; U, intron or 5'/3' UTR; single numbers refer to intergenic upstream or downstream mutations. The number of selective sweeps detected with G12 method and the group in which they are found are indicated in the last column.

Gene abbr.	Min. Temp. Winter	Min. Temp. Spring	Max. Temp. Spring	CV Prec. Spring	Prec. Spring	Prec. Summer	Germination rate 1	Flowering time	Sweeps G12
ACS7	1,2,3,4	1,2,3	1	5,6,7				S1,S2,8	1 *front*
NRT3.1	T1,N1	N1,N2		T1,N1,N2,N3, N4,N5,U1					1 *front*
PHYE	N6,U2	9				9,10,11,12,13,14,15,16,17			
DRP5A					T2,S3,U3,U4,U5,U6, U7,U8,U9,18,19				
NRT2.7						S4,S5,S6			1 *front*
ML5		9				N7,U10,U11,U12,U13,U14, U15,U16,U17,U18,9,14,15, 16,17			
MTN2							S7		

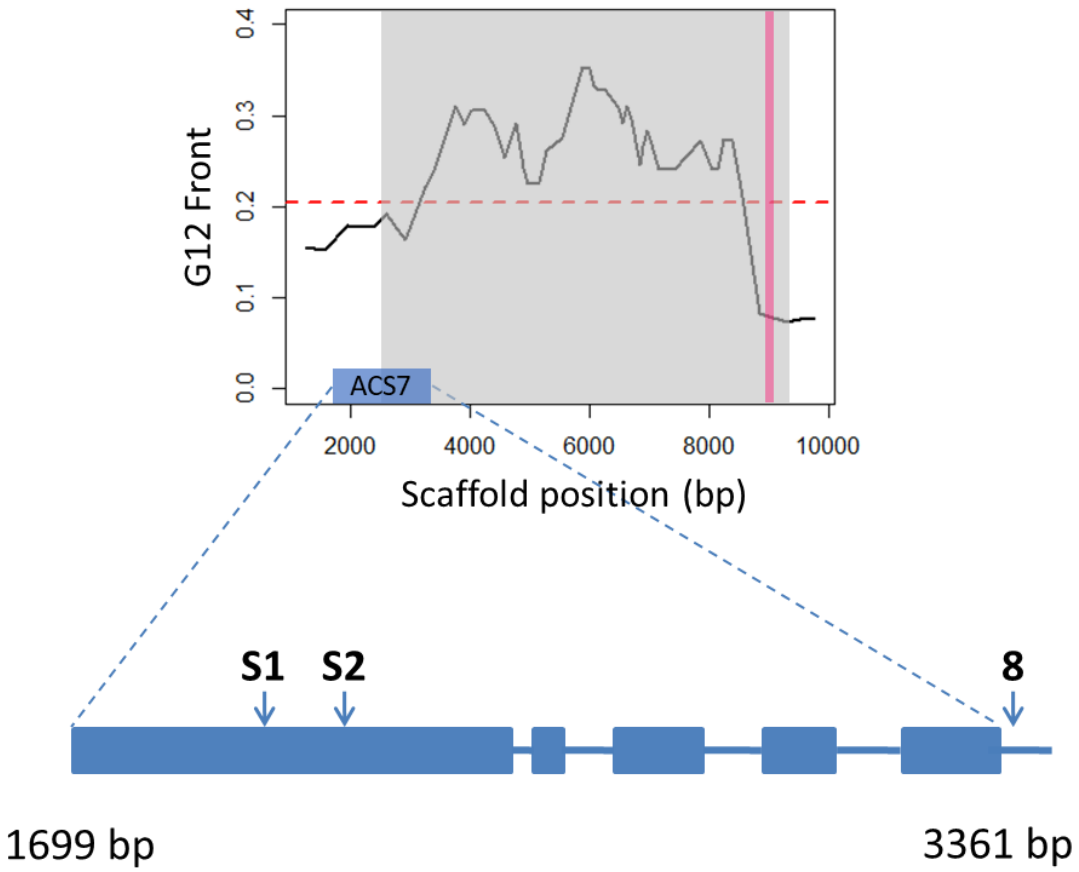


Figure 3.5. Selective signals detected in the putative *1-aminocyclopropane-1-carboxylate synthase 7* gene (ACS7). The gene is represented in blue, with five exons (bars) and four introns (lines) spanning from positions 1,699 bp to 3,361 bp of the scaffold 283046. The locations of the two outlier synonymous variants identified with BAYESCENV and BAYPASS are indicated as S1 (position 2,041 pb) and S2 (position 2,185 bp). A third outlier, located at position 3,391 bp, is indicated with an 8 (as in Table 3.2). The selective sweeps identified by G12 and SWEED methods in populations from the *front* range are shown in the graph. The grey area indicates a large significant sweep detected by G12 spanning from positions 2,503 to 9,388 bp. The black line corresponds to the G12 values obtained in the center of each genomic window with a size of 51 SNPs. The horizontal red dashed line shows the threshold for significant sweeps in G12 for *front* populations. The vertical red area corresponds to a significant sweep detected by SWEED around 9,000 bp.

To further investigate the link of ACS7 gene with the flowering time, we analyzed the relationship between the genotypes present at SNP positions S1 and S2 and the variation observed for this trait in the common garden. A Fisher's exact test revealed that both SNPs were tightly linked ($P < 0.001$), and most of the individuals genotyped (88.19%) carried one of these three specific combinations of alleles: homozygous for allele A at SNP positions S1 and S2 (AAAA, 17.72%), heterozygous AG at SNP positions S1 and S2 (AGAG, 29.96%) and homozygous for allele G at SNP positions S1

and S2 (GGGG, 40.51%). A clear geographic cline was evident in their distribution: while GGGG genotype was abundant in southern and central populations of the Iberian Peninsula (*core*), AAAA individuals were mainly found in populations of the expansion *front* (Fig. 3.6). Heterozygous AGAG genotypes were found across the whole sample, but tended to be more abundant in *front* populations (Fig. 3.6). The presence of AAAA genotypes was associated to an increase of the mean flowering time in northern populations (Fig. 3.6). The same trend was found in the close-relative outgroup, *L. saxatilis*, where the mean flowering time for genotypes carrying the combination AAAA (17 individuals) was 70.73 days, while genotypes AAAG (2 individuals) and AGAG (1 individual) invested only 49 and 55 days until the onset of the first flowering bud, respectively.

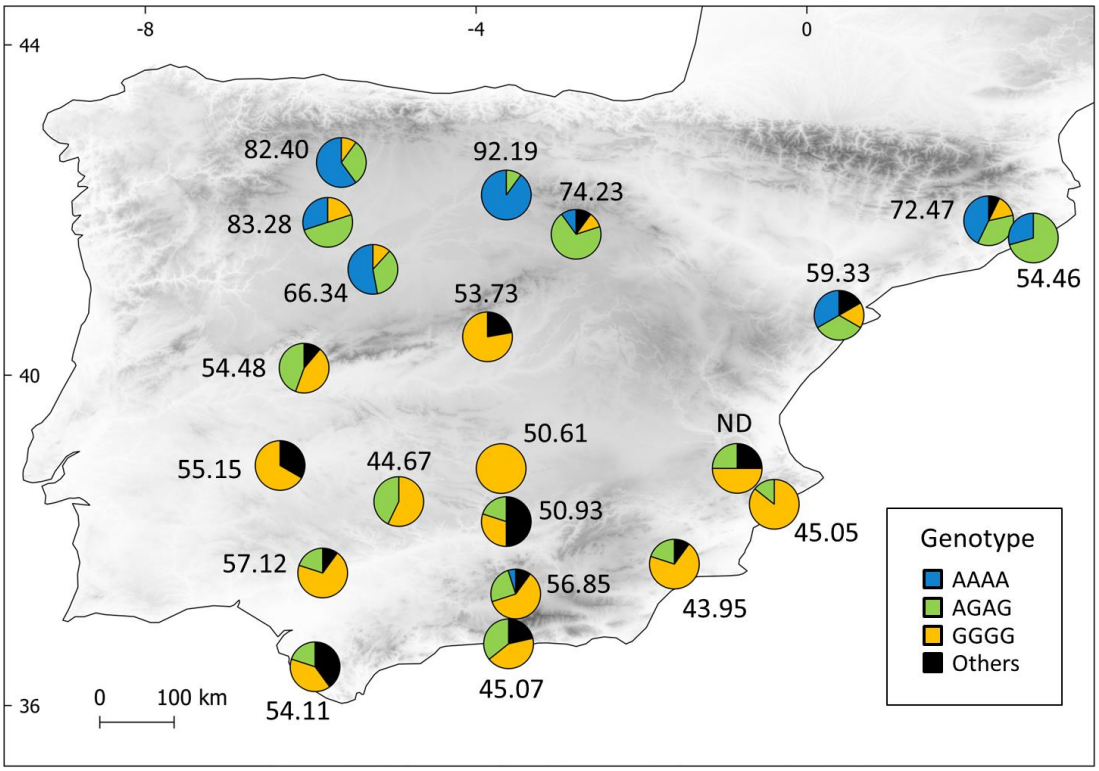


Figure 3.6. Geographic distribution of the two linked synonymous variants (SNPs S1 and S2) located on the first exon of the putative *1-aminocyclopropane-1-carboxylate synthase 7* gene (ACS7). Numbers close to pie charts refer to the mean flowering time for each population obtained in the late winter common garden experiment (retrieved from de Pedro et al., unpublished). ND = no data.

Given that our genotypic and phenotypic datasets were not based on the same sampling, we further explored this association restricting the analyses to the 78 genotyped plants that shared a half-sib individual (same mother-plant) with known phenotype grown in the common garden experiments. This subset included 15 populations distributed across the Iberian Peninsula (Table S3.4) that had similar proportions of the three main genotypes observed for the whole sampling: AAAA (15 individuals, 19.23%), AGAG (26 individuals, 33.33%) and GGGG (28, 35.90%). We used linear mixed models to test for the effect of genotype on flowering time, treating population as a random effect, using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2019). Maternal families carrying genotypes AAAA had significantly later flowering (70.3 days on average; $P < 0.05$) than those carrying heterozygous AGAG and homozygous GGGG combinations of alleles, which begun bolting around two weeks earlier (Fig. 3.7). The small flowering difference (56.6 vs. 55.4) between heterozygous AGAG and homozygous GGGG individuals was not significant.

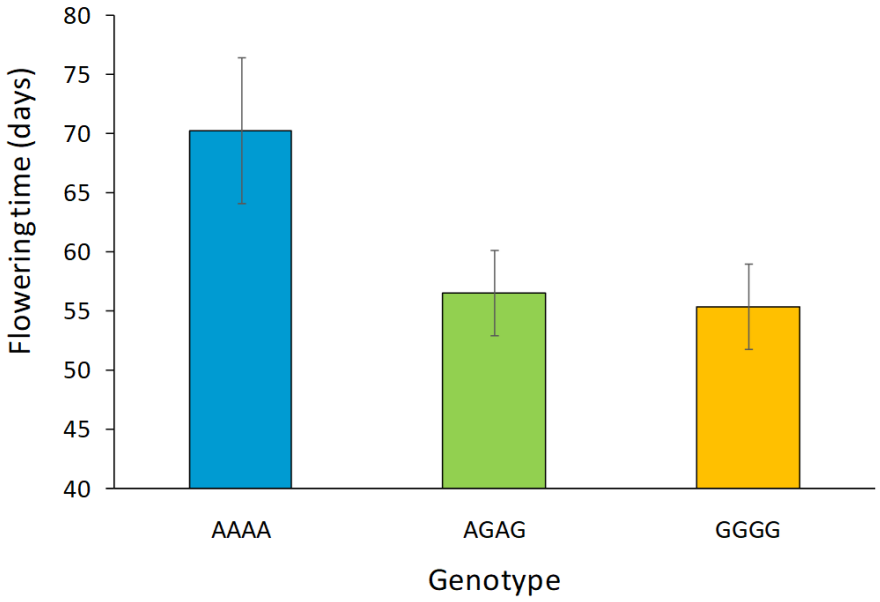


Figure 3.7. Mean flowering time (± SE) for individuals grown in a common environment sharing the same maternal family with plants genotyped as AAAA, AGAG and GGGG of the ACS7 gene (for further details, see text).

Genotype-environment associations, GEAs

Six candidate genes were involved in significant GEAs. The minimum temperatures of winter and spring, the precipitation during spring and summer, and the spring variation in precipitation were the climatic variables with higher number of significant outlier SNPs (Table 3.2), whereas no significant associations were found for the precipitation in autumn and winter.

One of the genes involved in significant GEAs was the putative *1-aminocyclopropane-1-carboxylate synthase 7* gene (ACS7), with seven upstream outlier SNPs associated with winter and spring temperatures, and/or variation in spring precipitation (Table 3.2). All these SNPs were tightly linked to each other, as well as to the abovementioned SNPs (S1, S2 and 8) associated with flowering time ($r > 0.95$).

The remaining five candidate genes had a variable number of SNPs linked to climate. Of particular interest were two putative nitrate transporters (Table 3.2), the *high affinity nitrate transporter 3.1* (NRT3.1) and the *high affinity nitrate transporter 2.7* (NRT2.7), which had signals of exclusive selective sweeps (G12 test) that were within the top 15 in *front* populations (Fig. 3.4). Both genes exhibited high G2/G1 values, supporting soft selective sweeps. In NRT3.1, the putative selective sweep affected a large intron located between the first two exons, while SNP outliers were essentially distributed in the third exon. Variants included one stop-gain, and five non-synonymous and one intron variants that were related to the variation of precipitation during spring, although some of them were also associated with minimum temperatures of winter and spring (Table 3.2). In NRT2.7, the selective sweep affected almost the entire gene, and the three outliers were synonymous mutations linked to summer precipitation (Table 3.2). Another relevant candidate gene involved in GEAs was a putative *phytochrome E* gene (PHYE), with one non-synonymous and one 3'UTR variant linked to the minimum temperature in winter, and nine upstream/downstream variants related to summer precipitation (Table 3.2). Two additional candidate genes, a *dynamin-related protein 5A* (DRP5A) and a *protein MEI2* (ML5), had 11 and 15 outlier SNPs associated to spring and summer precipitation, respectively (Table 3.2).

3.4 Discussion

Neutral and selective processes are expected to interact in a complex way during range expansions. At the expanding front, low genetic diversity and expansion load may limit adaptive potential. However, when species expand along environmental gradients, the speed of expansion is expected to be reduced by the need to adapt to the novel conditions, which can rescue populations from suffering accumulated expansion load (Gilbert et al., 2017). Because standing genetic variation (SGV) may be a major source of beneficial alleles in a new environment (e.g., Barrett & Schluter, 2008), we predicted a predominant role of adaptation from SGV in expanding populations. Here, we investigated these questions in the colonizing plant *L. longirostris*, a non-model Asteraceae which recently expanded its range in the Iberian Peninsula. Integrating distinct genomic approaches with phenotypic information obtained from common garden experiments, we found ample evidence for adaptation in the expanding front, occurring mainly from SGV, and identified a set of genes potentially underlying the changes in phenological traits that occurred along with the expansion, as discussed below.

Evolutionary dynamics at the expanding front

Despite the lower genetic diversity and expansion load characterizing populations from the *front*, we found evidence that adaptation has occurred during the *L. longirostris* northward expansion. Analogous climatic gradients to that of the present time characterized the Iberian Peninsula when the species expanded its range (e.g., González-Sampériz et al., 2010), which would have imposed new selective pressures as colonization advanced. The higher number of exclusive sweeps identified in *front* populations (79) relative to the *core* (49) likely reflects the progressive adaptation to these novel environments, and suggests that enough adaptive genetic variance was preserved during the expansion, at least for some phenotypic traits, as previously indicated by common garden experiments (de Pedro et al., unpublished). Although the potential for adaptation is often assimilated to the overall level of genetic variation, genetic diversity measured with neutral markers does not necessarily reflect the additive genetic variance present at ecologically relevant traits (Lewontin, 1965). Most of them have a highly polygenic basis, integrating the effect of multiple genes. The additive

variance for these traits could have been less affected by founder events than neutral genetic variation, because demographic bottlenecks mainly cause the loss of rare alleles at individual loci (Lewontin, 1965). Moreover, the lower genetic diversity detected in front populations could be in part reflecting the response to natural selection, which tends to reduce or even eliminate the molecular diversity around selected beneficial alleles (Maynard Smith & Haigh, 1974). Traits affected by few loci, by contrast, could have been more sensitive to sampling effects, but also could have responded more strongly to selection if they were governed by large-effect alleles that were beneficial in the novel environment at the expanding edge (Dlugosch et al., 2015).

The higher number of selective sweeps characterizing populations at the expanding front also points to a non-relevant effect of expansion load on the adaptive capacity of *L. longirostris*. Under certain circumstances, such as the presence of a strong environmental gradient, maladaptation can counteract the negative effect of expansion load (Gilbert et al., 2017). When species expand along environmental gradients, gene flow from the center of a species' range can cause local maladaptation in populations at the advancing front, which can eventually prevent the range from expanding outward (Kirkpatrick & Barton, 1997). This process can be beneficial to decrease expansion load at the range margins, because local maladaptation slow substantially the speed of the expansion, enabling fitter alleles to reach the edge and increase the efficacy of selection relative to drift (Gilbert et al., 2017). At the end, the frequency of deleterious alleles is reduced and only locally adapted populations are able to persist at the range edge (Gilbert et al., 2017).

According to these predictions, levels of expansion load in *L. longirostris* were smaller than those found in other plant species and, despite the higher proportion of deleterious to non-deleterious mutations found in *front* relative to *core* populations, harmful variants were not fixed at the population level (de Pedro et al., 2021). This suggests that the need to adapt to new environments could have contributed to reduce the severity of expansion load at the range margins. Alternatively, adaptation itself could be behind the accumulation of deleterious mutations in the expanding front, since the new selective pressures faced by colonizing individuals could have favored the spread of beneficial alleles, indirectly increasing the frequency of deleterious alleles because of hitchhiking with nearby selected variants (Hartfield & Otto, 2011).

Role of standing genetic variation (SGV) to adaptation

Our results revealed substantial differences regarding the predominant mode of selection in populations from the *core* and the *front*. In particular, we found a significant increase in the proportion of putative soft sweeps in *front* populations compared to the *core*, suggesting that standing variation had a prominent role for rapid adaptation to the novel environmental challenges associated with migration to northern habitats. However, identifying the traces of adaptation can be challenging, especially when populations are not in equilibrium, so we cannot discard that the excess of soft selection signatures in the front was due to the confounding signature of other processes. For instance, mild bottlenecks might have produced the stochastic fixation of several allele combinations, resulting in MLG structures similar to soft sweeps (Harris et al., 2018). If that was the case, however, we should expect distinct combinations to be randomly fixed in each population and, in *L. longirostris*, the first- and the second-most frequent MLGs characterizing soft sweeps were generally present in all populations (Fig. S3.1). On the other hand, soft selection can be easily inferred in the presence of "shoulders" of completed hard sweeps (Schrider et al., 2015). This "soft-shoulder" effect results from mutation and recombination events in the flanks of hard sweeps, which may be mis-characterized as being the target of a soft sweep. We could suspect such an effect if soft sweeps were especially frequent in the *core*, because the larger effective population sizes and older ages relative to the *front* could have favored higher rates of mutation and recombination, but this is not the case. Admixture between distinct populations can also generate MLG patterns that can mimic those produced by soft sweeps (Harris et al., 2018). In *L. longirostris*, high migration rates were inferred both within and between populations (de Pedro et al., 2021), which could have produced spurious signals of selective sweeps. Although this cannot be completely ruled out, in the presence of admixture we could expect a higher similarity of MLG patterns between closer populations (e.g., 68 and 72 on the one hand, and 67, 74, 78 on the other; see Fig. 3.1), which again does not seem to be the case (Fig. S3.1). Finally, background selection, whereby the elimination of deleterious mutations results in the removal of linked variants, is another potential confounding factor that may also spuriously resemble a sweep (Charlesworth et al., 1993). However, in the case of *L. longirostris* similar $\theta_{\pi N}/\theta_{\pi S}$ ratios have been reported throughout the species' range (de Pedro et al., 2021),

so similar background selection signatures should be expected in *core* and *front* populations.

Despite the excess of soft sweeps in the *front* could partially arise from some of the processes described above, recent studies show that "soft" patterns of adaptation are indeed common in a broad range of organisms (e.g., viruses, Pennings et al., 2014; fruit flies, Garud & Petrov, 2016; plants, Raquin et al., 2008; humans, Schrider & Kern, 2017), suggesting a key role of SGV to respond quickly to evolutionary challenges. The importance of SGV in rapid evolutionary processes is also supported by experimental studies with microbial communities (Gralka et al. 2016), mussels (Bitter et al., 2019) and sea urchins (Brennan et al., 2019). Although we cannot discard that some of the soft sweeps detected in *L. longirostris* were produced by recurrent *de novo* adaptive mutations, a predominant role of SGV as a source for adaptation in the expanding front of our study system is suggested by the fact that, for all the SNPs showing signals of selection, both SNP variants were present (MAF $\geq$ 5%) in the *core* and *front*, likely representing standing variation already present at the onset of the expansion. These results are consistent with the view that, compared with new mutations, SGV can facilitate evolution at shorter time scales because selection can act on alleles already present in the population at the onset of selection (Messer & Petrov, 2013), which can be especially important for the spread of colonizing species to novel environments.

In contrast to the *front*, we found a much higher frequency of hard sweeps in populations from the *core*. In agreement with our results, African (core) populations in *Drosophila melanogaster* showed evidence of hard and soft sweeps, whereas soft sweeps predominated in a recently expanded American population (Garud & Petrov, 2016). One possible explanation for these results could be that some demographic processes have altered the signal of soft sweeps in the *core* populations. For instance, population bottlenecks can result in the stochastic loss of some of the genetic background carrying the adaptive alleles, "hardening" initially soft selective sweeps (Wilson et al., 2014). A recent demographic decline was inferred for most populations of *L. longirostris* in the Iberian Peninsula (de Pedro et al., 2021), but it affected equally populations from southern and northern latitudes, so it can hardly explain the differences observed regarding the number of hard sweeps. Instead, our results are consistent with theoretical expectations that beneficial mutations are more frequently fixed in large populations (Crow & Kimura, 1970; Ohta, 1992). Populations from the

core were older and had larger *Ne* than those of the expanding *front* (Fig. 3.2; see also de Pedro et al., 2021), which could have provided more opportunities for the emergence of new beneficial mutations, and natural selection could be more effective relative to drift, favoring its spread throughout the population.

Candidate genes associated with range expansion

During the expansion, *L. longirostris* individuals have encountered novel environments and, thus, changes in resource availability, duration of favorable period for growth and reproduction, and biotic interactions. Our common garden experiments suggested important shifts in some phenotypic traits as *L. longirostris* expanded its range. Compared to the south, northern plants showed earlier germination and later flowering, resulting in longer growth cycles and, consequently, greater size and biomass (de Pedro et al., unpublished). This suggests that northern plants have evolved in response to increased water availability and milder temperatures, extending the favorable periods for growth.

Some of the candidate genes detected with our distinct approaches could potentially reflect these changes at the genetic level. For instance, the variation found in three of these candidates could be associated to the distinct germination behavior detected for southern (*core*) and northern (*front*) populations. On the one hand, we found a significant association between the germination rate in early fall (germination rate 1) and a synonymous variant located on the gene putatively encoding for *5-methylthioadenosine/S-adenosylhomocysteine nucleosidase 2* (MTN2). This gene encodes an essential enzyme within the methionine cycle (Yang cycle), a cycle that in plants plays an important role in sustaining ethylene production (Bürstenbinder et al., 2007). Ethylene, together with abscisic acid (ABA) and gibberellins (GAs), is part of the complex internal signaling network regulating dormancy and germination (Arc et al., 2013).

In addition to plant hormones, seed dormancy and germination are influenced by a variety of external factors such as temperature, water content, light conditions and nitrogen availability, which act as signals to determine whether the environmental conditions are suitable for germination and subsequent seedling emergence (see Yan & Chen, 2020 for a review). Accordingly, we found selective footprints in two additional

genes that could be reflecting adaptive responses to some of these environmental cues. One is the putative *high affinity nitrate transporter 2.7* (NRT2.7), which is a nitrate transporter that specifically controls nitrate content in seeds (Chopin et al., 2007). Seed nitrate, rather than a nutritional function for seed maturation, regulates the expression of the *CYP707A2* gene, which belongs to a small multigene family involved in ABA catabolism (Matakiadis et al., 2009). The level of nitrate accumulated in dry seeds negatively correlates with the ABA content and with the level of dormancy (Matakiadis et al., 2009). The other is the putative *phytochrome E* gene (PHYE), which is a red light-receptor with a prominent role in seed germination of *A. thaliana* at cool temperatures (Heschel et al., 2007).

Although we did not detect significant associations between these two genes and any of the variables related to germination included in this study (i.e., germination rates 1 and 2, germination time), the variation found for NRT2.7 gene and in the regulatory regions of PHYE was significantly linked to summer precipitation (Table 3.2), a climatic variable that was also related to different germination behaviors detected under a common environment (de Pedro et al., unpublished). Whereas germination of seeds from drier sites was mostly restricted to late winter conditions, seeds from populations with higher precipitation in summer were able to germinate both in early fall and late winter, suggesting that dormancy could be relaxed in more humid environments (de Pedro et al., unpublished). The importance of summer precipitation to explain variation in seed dormancy has also been reported for populations of *A. thaliana* at the European scale, with populations receiving more precipitation in the summer being less dormant (Kronholm et al., 2012). The fact that similar climate associations were found for the germination behavior and for genes acting in hormonal (MTN2) and signaling pathways (NRT2.7 and PHYE), suggests an integrated adaptive response to optimize one of the most important life-cycle events (germination) to the environmental conditions, in particular those related to water availability.

In the same way, a coordinated response to the environment could be behind the variation found for two genes potentially related to flowering. At the phenotypic level, the variation found for this trait was mostly correlated with minimum temperatures of spring and summer (de Pedro et al., unpublished). Similarly, there was a significant association with the minimum temperature of winter and the variation present in the exonic regions of PHYE, a gene that is particularly important for the control of

flowering under cool temperatures (Halliday & Whitelam, 2003). Significant associations with minimum temperatures of winter and spring were also detected for several SNPs located in the upstream region of the putative *1-aminocyclopropane-1-carboxylate synthase 7* (ACS7) gene, which belongs to a multigene family encoding a group of enzymes catalyzing the conversion of S-adenosyl-methionine (SAM) into 1-aminocyclopropane-1-carboxylate (ACC), the direct precursor of ethylene (see Pattyn et al., 2021 for a recent review). Different ACS isoforms interact in a complex way to regulate multiple ethylene-mediated processes, including flowering time (Tsuchisaka et al., 2009).

Interestingly, this was one of the best candidate genes for selection in *L. longirostris*, showing selective footprints with the three distinct approaches (G12, SweeD and association methods). The significant relationship between two tightly linked synonymous outliers located on the first exon of ACS7 and the flowering time measured in the common garden strongly supports its link with this phenotypic trait. Synonymous mutations have traditionally been considered to be silent with respect to fitness, but there is growing evidence that they affect the expression and function of the translated protein and, therefore, are under selective pressure (Hunt et al., 2014). Delayed flowering was associated with a particular genotype (AAAA) that was present almost exclusively in the populations of the expansion front. The same genotypic combination was common in the sister species *L. saxatilis*, which is also characterized by late flowering compared to *L. longirostris*. Although we cannot completely rule out that the environmental associations detected for the regulatory region of ACS7 were in response to processes other than flowering, SNPs were associated with very similar climatic drivers than those explaining the variation in flowering time at the phenotypic level, suggesting their involvement in the evolution of this trait.

Finally, one of the strong candidates for selection detected with both CLR and G12 tests was the putative *enhanced disease resistance 2* (EDR2) gene, which was exclusive of the populations in the range *front*. In Arabidopsis, EDR2 acts as a negative regulator of the salicylic acid-mediated resistance to pathogens (Tang et al., 2005). In the common milkweed (*Asclepias syriaca*), changes in inducible defense responses, including an increase of production of salicylic acid after herbivore attack, have been found for introduced populations compared to their native counterparts (Agrawal et al., 2015). This is consistent with life-history theory predicting that, during the course of

expansions, specific defenses tailored to deal with coevolved or specialized enemies are expected to be replaced by a higher investment against generalist enemies (Phillips et al., 2010). Our results point to the same direction, and suggest that the selective sweep affecting EDR2 in *L. longirostris* is likely reflecting evolved defense responses to altered pathogen pressures during the colonization into novel environments.

Concluding remarks

Our analysis confirmed that adaptation to new environments has occurred during the northward expansion of *L. longirostris* in the Iberian Peninsula, as suggested previously by the phenotypic variation found for several life-history traits. Adaptation has occurred despite the decrease in genetic diversity and the increase in expansion load characterizing populations of the range front, suggesting that the loss of genetic diversity probably affected non-relevant (neutral) loci and that expansion load could be the result of hitchhiking of deleterious alleles with nearby positively selected variants. We also found ample evidence that adaptation during the expansion mainly proceeded by selection acting on genetic variation already present in the populations (i.e., SGV), highlighting the importance to preserve species' genetic variation as a source for future adaptation to anthropogenic climatic change. We also identified several candidate genes, mainly acting in signaling pathways, potentially mediating the phenological changes detected at the phenotypic level. Two of them (ACS7 and MTN2) were part of the ethylene signaling pathway and the adjacent Yang cycle, thus representing good candidates for further analyses. Other strong candidates were involved in the integration of external light (PHYE) and nitrate (NRT2.7) signals into the germination and flowering pathways. Summer precipitation and minimum temperatures of winter and spring were detected as key environmental cues driving the variation found in genes putatively involved in germination and flowering, respectively, confirming previous results obtained at the phenotypic level. Finally, the selective signals found for another gene (EDR2) suggest that altered defense responses could have evolved during the course of expansion.

General Discussion

Unraveling the interaction of the different ecological and evolutionary processes governing the limits of species' geographical ranges is essential to understand why some species fail to expand while others easily spread across different habitats. This is especially important under the unprecedented rates of climate change, which are expected to drive important range shifts in many organisms. Although historically ecological factors have received more attention, the study of the evolutionary processes leading to or constraining range expansions and shifts has recently become a major focus of both theoretical and empirical studies.

Recent theory predicts that adaptation can be limited when species expand their ranges due to several demographic and genetic processes, such as the loss of genetic diversity (Slatkin & Excoffier, 2012), the accumulation of deleterious mutations by gene surfing (e.g., expansion load, Peischl et al., 2013) and/or extensive gene flow from core populations (Kirkpatrick & Barton, 1997). In addition, evolution at range margins can be constrained by correlation among traits that are subject to antagonist selection pressures (Etterson & Shaw, 2001).

The few empirical plant studies exploring the patterns and evolutionary consequences of range expansions are mostly restricted to the Brasicaceae family, e.g. the model plant *Arabidopsis thaliana* (Williams et al., 2016) and its relatives (Laenen et al., 2018; Wang et al., 2018; Willi et al., 2018; Kryvokhyzha et al., 2019; Takou et al., 2019a). In order to gain some generality, we explored the demographic and evolutionary dynamics of range expansion in the non-model species *Leontodon longirostris*, a short-lived colonizing Asteraceae with a different demographic history and life-history traits.

Contrary to the prevailing pattern of multiple refugia that is commonly found for many plants in the Iberian Peninsula ("refugia-within-refugia" model, Gómez & Lunt, 2007), our results suggested that the overall demographic history of the species was dominated by a progressive south-north expansion starting around 40,000 yrs ago. In accordance to theoretical and empirical evidence, we found that the range expansion was accompanied by a loss of genetic diversity and a significant increase in the proportion of putatively deleterious mutations (expansion load) in the range-front. Nevertheless, levels of expansion load were smaller than those reported for other plant species (e.g., González-Martínez et al., 2017; Laenen et al., 2018), since none of the deleterious variants detected was fixed (e.g., present in all the individuals) in the front populations. In addition, we found ample evidence suggesting that adaptation to new

environments has occurred during the northward expansion. This is supported by the phenotypic variation found for key life-history traits, which strongly support the presence of an adaptive cline for major life-history stage transitions. Similarly, the high number of exclusive sweeps identified in front populations, together with the significant association of genetic variants (SNPs) with environmental and phenotypic variables, also indicates that enough genetic variance was preserved during the expansion for selection to act efficiently. It is also remarkable that our results revealed that standing genetic variation was the predominant source for adaptation during *L. longirostris* northward expansion, in accordance with the claim that genetic variation already present in the populations provides the potential for rapid evolution under new environments (Barrett & Schluter, 2008; Messer & Petrov, 2013). These findings provide important insights for understanding adaptation and conservation of species in the Anthropocene.

There are several species-specific traits and environmental features that could explain these results. On the one hand, the species has maintained a self-incompatible mating system along the expansion route, which has most likely contributed to reduce the severity of bottlenecks, lowering the role of genetic drift and gene surfing. While selfing is generally expected to evolve in colonizing annual plants (Baker, 1955), the maintenance of self-incompatibility could have been facilitated by the high dispersal capacity of the species, which can compensate mate limitation on the leading edge, in agreement with theoretical models (Pannell & Barrett, 1998). In contrast, plants evolving towards selfing during range expansions have been associated to stronger levels of expansion load (e.g., Laenen et al., 2018; Willi et al., 2018; Koski et al., 2019). On the other hand, *L. longirostris* is able to establish temporary seed banks (Ruiz de Clavijo, 2001), which could have also contributed to maintain enough additive genetic variance for ecologically relevant traits, and thus further favoring selection over drift.

Besides intrinsic factors, recent theory predicts a reduction of expansion load when species expand along steep environmental gradients (Gilbert et al., 2017). This is expected because steep environmental gradients slow down the rate of expansion by the need of range-front populations to adapt to the novel conditions, limiting the mutational burden associated to range expansion (Gilbert et al., 2017). Accordingly, we found that the northward expansion has been a slow process, taking around 15,000 years. It must be noted, however, that most deleterious mutations that were found in *L. longirostris* affected splice sites, and then were located at the exon-intron boundaries. This fact

points to an alternative explanation, suggesting that strong positive selection acting on coding regions could have produced the hitchhiking of nearby deleterious alleles in populations of the expansion front (Hartfield & Otto, 2011).

The adaptive cline found for major life-history stage transitions was consistent with life-history theory predicting distinct adaptive strategies linked to stressful and/or predictable environments (e.g., Schaffer, 1974; Stearns, 1992; Stearns et al., 2001). Our results also confirmed the adaptive relevance of germination and flowering traits to optimize plant performance to local environmental conditions. Both traits had a strong effect, directly or indirectly, on final plant size and biomass, supporting that phenological patterns jointly determined overall plant fitness. This is consistent with the view that combinations of developmental traits often evolve as trait syndromes, as has been suggested for the model plant *Arabidopsis thaliana* (Takou et al., 2019b). This was further supported by our results at the molecular level, where the variation found in several genes, potentially underpinning phenological traits, suggested a coordinated adaptive response to optimize life-cycle events to similar environmental drivers. Furthermore, despite the lack of phylogenetic relatedness and contrasting demographic histories, the patterns of selection found in this study strongly mirrored those reported for the annual model plant *A. thaliana* in the same area (Manzano-Piedras et al., 2014; Vidigal et al., 2016), suggesting that similar ecological strategies have evolved in response to regional differences in environmental conditions.

Overall, then, we can clearly infer that the current observed levels of expansion load do not seem to have prevented local adaptation to the environmental variation encountered during expansion. However, we are aware that additional and stronger experimental evidence on adaptation could be provided to rigorously test that hypothesis, for instance using reciprocal transplant experiments between core and front populations (Kawecki & Ebert, 2004). Similarly, our results suggest that processes actually constraining the current geographical range of the species could probably be associated to other evolutionary constraints, such as limited genetic diversity and/or genetic correlations among traits (see, for instance, Kalisz & Kramer, 2008). Testing for the relative importance all these factors constraining the current species range could also be experimentally performed by transplanting populations with different combinations of genetic variability, expansion load, and varying types and strengths of genetic correlations between key life-history traits, beyond its current the range limits.

Conclusions

From this work, the following main conclusions can be drawn:

1. Contrary to the prevailing pattern of multiple refugia commonly found for many plants in the Iberian Peninsula, we did not find divergent lineages for *L. longirostris*. Instead, the demographic history of the species revealed a progressive south to north expansion starting around 40,000 yrs BP.
2. In agreement with theoretical predictions, the range expansion was accompanied by a loss of genetic diversity and a significant increase in the proportion of putatively deleterious mutations (expansion load) in range-front populations.
3. Levels of expansion load in *L. longirostris* were smaller than those found in other plant species. The high dispersal ability of the species, preventing the evolution of selfing in the expanding front, could explain these results.
4. We found ample phenotypic and genomic evidence indicating that adaptation has occurred during the expansion process. These results give support to theoretical predictions that expansions along steep environmental clines can reduce expansion load; alternatively, expansion load in front populations could also be an indirect effect of adaptation, due to the hitchhiking of deleterious alleles along with positively selected variants.
5. Our results revealed a predominant role of standing genetic variation as a source for selection during *L. longirostris* northward expansion, supporting the view that variation already present in the populations provides the potential for rapid evolution under new environments.
6. Consistent with the recent view that multiple traits evolve in combination as trait syndromes, both the timing of germination and flowering had a strong effect on other traits such as lifespan, size at reproduction and total biomass. This was further supported by our results at the molecular level, suggesting a coordinated adaptive response of several candidate genes -potentially underlying phenological traits- to optimize life-cycle events to a given environment.
7. Most of the candidate genes identified were involved in signaling pathways, suggesting an important role to adaptation of genes directly affected by environmental stimuli.

8. Summer precipitation and temperature seem to be the key environmental cues driving the variation in phenological and molecular patterns of variation involved in adaptation.
9. This study shows that the non-model, short lived and colonizing species *L. longirostris* is a promising alternative species to gain a new insight into the adaptive dynamics of range margins, complementing current knowledge limited to a few model species.

— Supplementary Material —

Chapter 1

Figure S1.1. Proportion of gene pool membership by population inferred from 168,733 SNPs with FASTSTRUCTURE (*K* from 2 to 5). Pie charts show averaged values of three different runs.

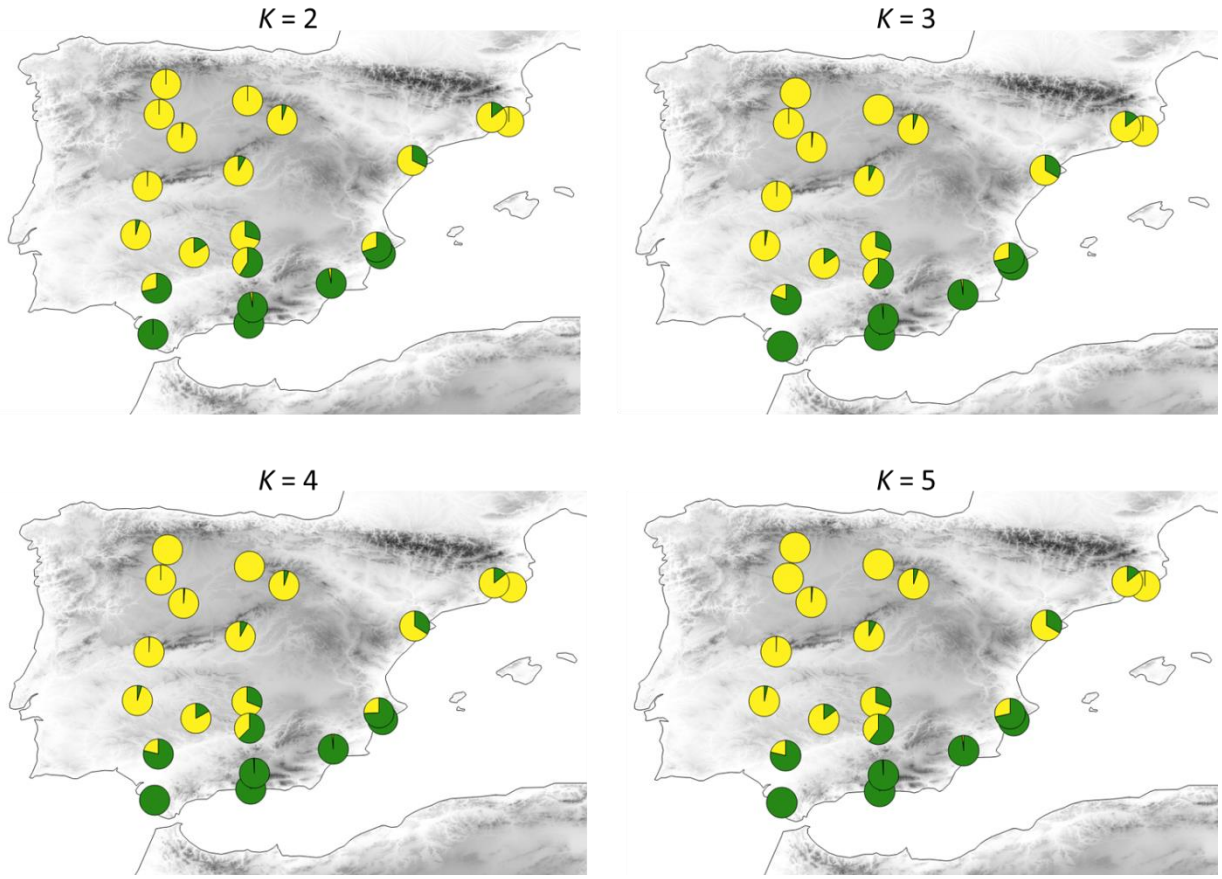


Figure S1.2. Third and fourth axes of the principal component analysis (PCA) of the genetic variation found in 238 individuals of *Leontodon longirostris* in the Iberian Peninsula (based on 168,733 SNPs). Each individual is represented by a colored point. Colors are the same that in Fig. 1.3 of the main text. Population details can be found in Tables 1.1 and S1.1.

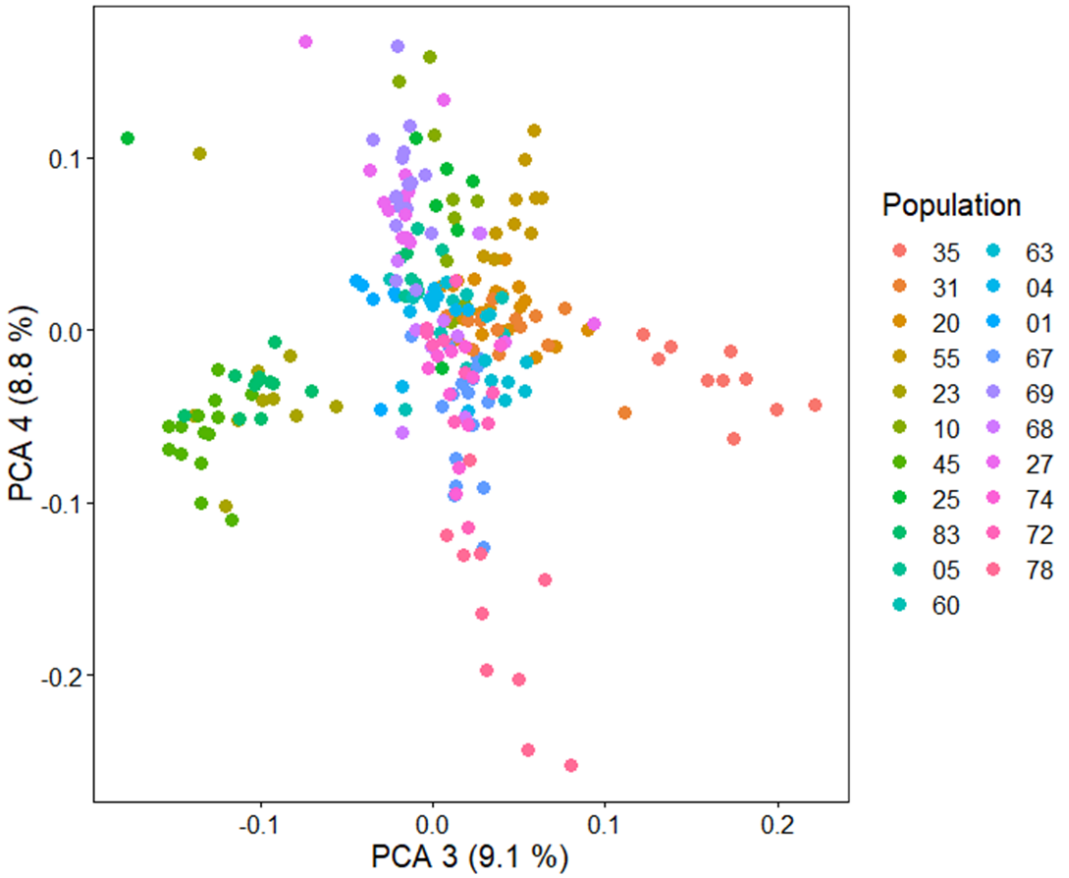


Figure S1.3. Relationship between geographical (logarithmic) and genetic distances [F_{ST}/(1- F_{ST})] among *Leontodon longirostris* populations (based on 116,946 SNPs with known allelic state). Mantel test: $r = 0.56$, $P < 0.001$.

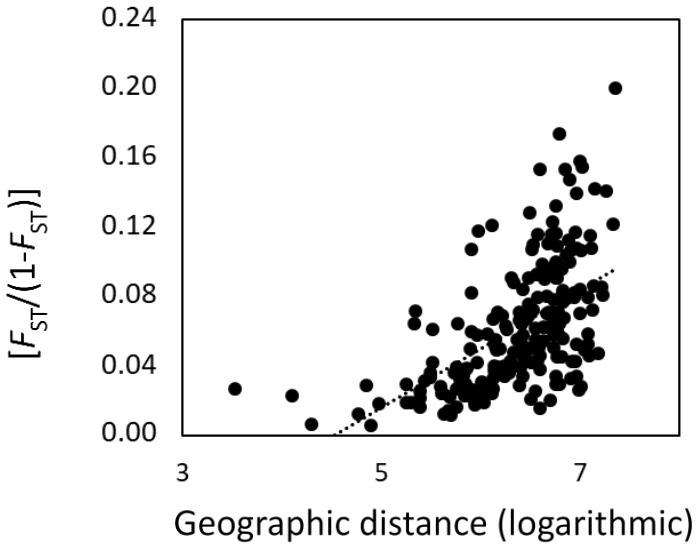


Figure S1.4. Estimates of *Leontodon longirostris* range expansion based on the pairwise directionality index (ψ). Colors show directionality from most likely (yellow) to unlikely (red and grey). Sample locations are represented by circles, with colors measured as heterozygosities (light circles indicate high heterozygosity, dark circles indicate low heterozygosity). The inferred source is indicated with "X" ($r^2 = 0.34$, $P < 0.001$).

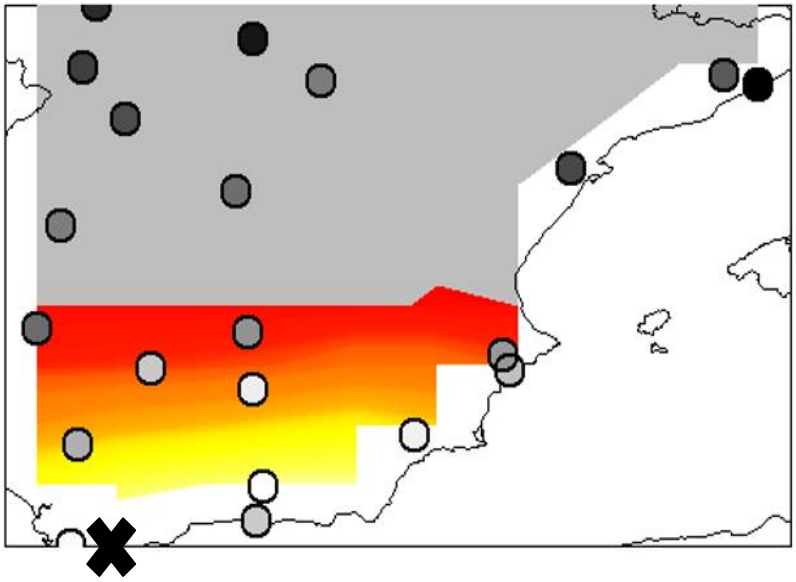


Table S1.1. Name and geographical location of the 21 populations of *Leontodon longirostris* sampled in the Iberian Peninsula.

Population code	Population name	Province	Latitude (º)	Longitude (º)
01	Mas de Barberans	Tarragona	40.72	0.39
04	Mahadahonda	Madrid	40.46	-3.86
05	Almagro	Ciudad Real	38.87	-3.70
10	Carboneros	Jaén	38.23	-3.63
20	Monachil	Granada	37.14	-3.51
23	La Hoya	Murcia	37.71	-1.60
25	Dos Torres	Córdoba	38.47	-4.93
27	Tagamanent	Barcelona	41.78	2.33
31	Salobreña	Granada	36.75	-3.61
35	Medina-Sidonia	Cádiz	36.47	-5.95
45	El Campello	Alicante	38.44	-0.39
55	Cantillana	Sevilla	37.60	-5.85
60	Mérida	Cáceres	38.91	-6.37
63	Plasencia	Cáceres	40.08	-6.08
67	Alaejos	Valladolid	41.28	-5.24
68	Calatañazor	Soria	41.71	-2.79
69	Malgrat de Mar	Barcelona	41.66	2.74
72	Hontoria de la Cantera	Burgos	42.18	-3.63
74	Santa Eulalia de Tábara	Zamora	41.85	-5.79
78	La Virgen del Camino	León	42.57	-5.63
83	Coll de la Carrasqueta	Alicante	38.60	-0.49

Table S1.2. Number and proportion of SNPs in genic and intergenic regions of the genome of *Leontodon longirostris* (based on 238 individuals).

	SNPs (*N*)	**SNPs (%)**
Total	168,733	100.00
Intergenic	106,179	62.93
Genic	62,554	37.07
Coding (exons)	21,909	12.98
Non-coding (introns and UTR)	40,645	24.09

Table S1.3. Prior parameter distributions for the four demographic models tested with FASTSIMCOAL2. Mutation rate is the number of mutations per locus per generation. Times are given in number of generations, which can be translated directly to years in annual *Leontodon longirostris*, ignoring the effects of a possible seedbank.

Parameter	Description	Probability distribution	Minimum value	Maximum value
Standard model (Model 1 in Fig. 1.2)				
MUT RATE	Mutation rate	Uniform	1×10^{-7}	1×10^{-9}
N_e CUR	Current effective population size	Uniform	1	1×10^{5}
Three-epoch models (Models 2, 3, 4 in Fig. 1.2)				
N_e ANC	Ancestral effective population size	Uniform	1	1×10^{5}
N_e INT	Effective population size between ancestral and current periods	Uniform	1	1×10^{5}
N_e CUR	Current effective population size	Uniform	1	1×10^{5}
T ANC	Ancestral time of population size change	Uniform	1	1×10^{5}
T REC	Recent time of population size change	Uniform	1	1×10^{5}

Table S1.4. Pairwise F_{ST} values between the 21 populations of *Leontodon longirostris* included in this study. Values are based on the 116,946 SNPs with known allelic state. All values were significantly different from zero ($P < 0.001$) except those highlighted in grey.

	01	**04**	**05**	**10**	**20**	**23**	**25**	**27**	**31**	**45**	**60**	**63**	**67**	**68**	**69**	**83**	**35**	**55**	**72**	**74**
04	0.028																			
05	0.020	0.014																		
10	0.016	0.023	0.012																	
20	0.029	0.050	0.038	0.018																
23	0.044	0.066	0.057	0.035	0.022															
25	0.033	0.025	0.018	0.028	0.060	0.065														
27	0.034	0.028	0.032	0.033	0.055	0.077	0.050													
31	0.053	0.067	0.055	0.027	0.007	0.023	0.076	0.068												
45	0.055	0.078	0.066	0.049	0.034	0.016	0.083	0.085	0.039											
60	0.041	0.024	0.018	0.029	0.058	0.073	0.025	0.045	0.081	0.083										
63	0.043	0.016	0.023	0.038	0.064	0.082	0.026	0.044	0.085	0.091	0.021									
67	0.066	0.035	0.048	0.062	0.087	0.109	0.059	0.041	0.104	0.117	0.051	0.040								
68	0.047	0.024	0.035	0.047	0.073	0.090	0.048	0.025	0.087	0.099	0.042	0.041	0.036							
69	0.097	0.063	0.074	0.078	0.103	0.133	0.098	0.022	0.124	0.133	0.079	0.079	0.066	0.059						
83	0.030	0.053	0.036	0.032	0.033	0.031	0.063	0.058	0.046	0.026	0.055	0.064	0.086	0.070	0.104					
35	0.074	0.099	0.083	0.052	0.025	0.040	0.105	0.108	0.028	0.055	0.108	0.114	0.148	0.122	0.167	0.066				
55	0.050	0.062	0.040	0.034	0.037	0.055	0.060	0.075	0.048	0.067	0.058	0.063	0.097	0.083	0.123	0.059	0.067			
72	0.060	0.026	0.038	0.048	0.076	0.099	0.052	0.020	0.096	0.104	0.041	0.039	0.024	0.018	0.043	0.074	0.136	0.089		
74	0.056	0.022	0.033	0.044	0.072	0.094	0.047	0.027	0.091	0.101	0.037	0.031	0.028	0.024	0.048	0.073	0.128	0.084	0.012	
78	0.061	0.026	0.037	0.048	0.076	0.098	0.052	0.026	0.096	0.105	0.041	0.038	0.033	0.019	0.047	0.074	0.134	0.088	0.012	0.006

Table S1.5. Parameters inferred for the four demographic scenarios tested in FASTSIMCOAL2 for each population separately. The best-fitting demographic model is indicated in bold and the confidence intervals (95%) are within brackets. Times are given in number of generations, which can be translated directly to years in annual *Leontodon longirostris*, ignoring the effects of a possible seedbank. *Ne* ANC: ancestral effective population size; *Ne* INT: effective population size between ancestral and current periods; *Ne* CUR: current effective population size; T ANC: ancestral time of population size change; T REC: recent time of population size change; $Log_{10}L$: maximum likelihood estimate of the model; AIC: Akaike's Information Criterion value; ΔAIC: difference in AIC value from that of the strongest model; w_iAIC: Akaike's weight of evidence. Each specific parameter is illustrated in Fig. 1.2.

Core									
Demographic model	**N_e ANC**	**N_e INT**	**N_e CUR**	**T ANC**	**T REC**	**$Log_{10}L$**	**AIC**	**ΔAIC**	**w_iAIC**
Population 10									
(1) Standard (Stable)	-	-	38,337	-	-	-83,635.17	385,158	18,797	0.00
(2) Bottleneck + Expansion	20,603	230	49,691	76,176	29,114	-79,552.40	366,362	1	0.29
(3) Expansion + Bottleneck	256	61,395	49,825	29,186	26,387	-79,552.80	366,364	3	0.11
(4) No parameter restriction (Expansion)	**139** (54-7,819)	**4,378** (3,670-55,916)	**50,074** (48,930-52,170)	**29,540** (28,834-33,909)	**27,431** (11,989-28,135)	**-79,552.08**	**366,361**	**0**	**0.60**
Population 20									
(1) Standard (Stable)	-	-	50,166	-	-	-122,428.20	563,807	27,763	0.00
(2) Bottleneck + Expansion	44,297	2,402	60,535	67,648	24,042	-116,510.16	536,559	516	0.00
(3) Expansion + Bottleneck	2,389	88,103	60,922	23,878	23,361	-116,510.89	536,562	519	0.00
(4) No parameter restriction (Expansion)	**587** (219-861)	**24,374** (20,172-27,724)	**74,663** (71,780-77,735)	**29,893** (29,003-31,101)	**10,869** (9,280-12,480)	**-116,398.20**	**536,044**	**0**	**1.00**

Population 23									
(1) Standard (Stable)	-	-	29,062	-	-	-78,915.56	363,424	12,598	0.00
(2) Bottleneck + Expansion	413,01	459	37,702	87,002	30,884	-76,204.65	350,945	120	0.00
(3) Expansion + Bottleneck	589	41,932	38,041	30,308	28,612	-76,205.28	350,948	123	0.00
(4) No parameter restriction (Expansion)	**32** (4-255)	**30,659** (27,687-32,110)	**66,684** (47,371-87,335)	**31,787** (31,215-32,362)	**3,209** (2,322-7,692)	**-76,178.55**	**350,825**	**0**	**1.00**
Population 25									
(1) Standard (Stable)	-	-	34,715	-	-	-65,787.39	302,966	13,776	0.00
(2) Bottleneck + Expansion	31,719	985	47,849	82,218	26,313	-62,819.12	289,303	113	0.00
(3) Expansion + Bottleneck	925	60,515	47,344	26,587	25,145	-62,819.75	289,306	115	0.00
(4) No parameter restriction (Expansion)	**127** (9-397)	**24,433** (18,663-30,846)	**64,308** (58,205-79,076)	**28,825** (27,962-29,435)	**10,110** (6,382-13,949)	**-62,794.68**	**289,190**	**0**	**1.00**
Population 31									
(1) Standard (Stable)	-	-	17,357	-	-	-91,152.51	419,777	18,475	0.00
(2) Bottleneck + Expansion	42,995	1,682	46,761	82,354	23,248	-87,199.27	401,577	276	0.00
(3) Expansion + Bottleneck	1,682	62,737	46,566	23,226	22,590	-87,200.32	401,582	281	0.00
(4) No parameter restriction (Expansion)	**63** (9-406)	**16,503** (13,433-21,930)	**54,209** (51,982-56,862)	**28,020** (26,933-28,515)	**12,764** (10,281-14,704)	**-87,139.34**	**401,301**	**0**	**1.00**
Population 35									
(1) Standard (Stable)	-	-	44,289	-	-	-82,838.99	381,492	13,802	0.00
(2) Bottleneck + Expansion	70,807	3,140	48,073	86,619	23,842	-79,858.80	367,773	84	0.00

(3) Expansion + Bottleneck	3,209	48,444	46,959	23,797	717	-79,860.57	367,782	92	0.00
(4) No parameter restriction (Expansion)	**2,371** (1,643-2,769)	**24,699** (15,183-32,925)	**56,414** (53,308-63,466)	**27,144** (25,746-30,105)	**11,010** (6,599-15,197)	**-79,840.52**	**367,689**	**0**	**1.00**
Population 45									
(1) Standard (Stable)	-	-	25,068	-	-	-84,600.76	389,605	12,324	0.00
(2) Bottleneck + Expansion	27,570	1,222	34,715	54,762	26,006	-81,947.42	377,392	111	0.00
(3) Expansion + Bottleneck	1,026	53,058	34,124	26,410	25,753	-81,948.85	377,398	118	0.00
(4) No parameter restriction (Expansion)	**40** (8-596)	**18,536** (15,898-25,140)	**37,935** (36,702-39,914)	**29,504** (27,745-29,796)	**12,507** (7,514-14,908)	**-81,923.26**	**377,281**	**0**	**1.00**
Population 55									
(1) Standard (Stable)	-	-	16,619	-	-	-70,183.45	323,211	10,330	0.00
(2) Bottleneck + Expansion	11,965	776	32,777	64,387	26,321	-67,963.54	312,994	112	0.00
(3) Expansion + Bottleneck	1,096	41,229	10,281	24,870	580	-67,939.32	312,882	1	0.39
(4) No parameter restriction (Expansion + Bottleneck)	**1,168** (986-1,365)	**42,571** (40,153-54,632)	**9,990** (2,278-24,639)	**24,862** (24,136-25,436)	**626** (91-4,588)	**-67,939.12**	**312,881**	**0**	**0.61**

Intermediate									
Demographic model	N_e **ANC**	N_e **INT**	N_e **CUR**	**T ANC**	**T REC**	**Log$_{10}$L**	**AIC**	**ΔAIC**	w_i**AIC**
Population 04									
(1) Standard (Stable)	-	-	11,405	-	-	-63,444.53	292,177	13,993	0,00
(2) Bottleneck + Expansion	34,054	330	39,286	57,181	21,889	-60,406.52	278,192	8	0,01
(3) Expansion + Bottleneck	417	46,153	33,061	21,511	3,440	-60,405.60	278,188	4	0,12
(4) No parameter restriction (Expansion + Bottleneck)	**358** (258-508)	**42,506** (40,534-62,941)	**32,822** (27,539-38,983)	**21,628** (21,306-22,065)	**2,107** (808-13,782)	**-60,404.73**	**278,184**	**0**	**0.87**
Population 05									
(1) Standard (Stable)	-	-	17,190	-	-	-69,638.88	320,703	16,157	0,00
(2) Bottleneck + Expansion	41,958	769	47,061	48,058	22,811	-66,155.73	304,668	123	0,00
(3) Expansion + Bottleneck	854	46,485	42,052	22,692	25	-66,155.98	304,670	124	0,00
(4) No parameter restriction (Expansion)	**9** (3-321)	**12,461** (8,977-21,992)	**52,506** (50,459-56,382)	**25,531** (24,369-25,639)	**15,568** (10,570-17,397)	**-66,129.06**	**304,546**	**0**	**1.00**
Population 60									
(1) Standard (Stable)	-	-	48,733	-	-	-60,971.42	280,788	12,397	0.00
(2) Bottleneck + Expansion	32,296	14	34,482	61,132	22,500	-58,280.63	268,402	11	0.00
(3) Expansion + Bottleneck	73	36,756	25,017	22,492	964	-58,278.53	268,393	2	0.32
(4) No parameter restriction (Expansion + Bottleneck)	**98** (24-	**37,711** (36,505-	**19,717** (13,746-	**22,269** (21,920-	**630** (313-	**-58,278.20**	**268,391**	**0**	**0.68**

	185)	55,403)	33,705)	22,649)	11,958)				
Population 63									
(1) Standard (Stable)	-	-	13,155	-	-	-65,045.99	299,552	13,398	0.00
(2) Bottleneck + Expansion	12,031	409	38,468	77,279	23,120	-62,144.29	286,195	42	0.00
(3) Expansion + Bottleneck	473	39,758	38,982	22,917	19,183	-62,144.45	286,196	42	0.00
(4) No parameter restriction (Expansion)	**21** (3-240)	**16,294** (8,782-28,457)	**41,617** (40,208-45,328)	**24,286** (23,596-24,653)	**14,803** (7,664-18,383)	**-62,135.25**	**286,153**	**0**	**1.00**
Population 68									
(1) Standard (Stable)	-	-	17,654	-	-	-63,772.33	293,686	10,275	0.00
(2) Bottleneck + Expansion	16,358	8	29,814	74,334	24,453	-61,568.17	283,542	131	0.00
(3) Expansion + Bottleneck	133	47,955	21,576	24,047	4,900	-61,541.24	283,418	6	0.04
(4) No parameter restriction (Expansion + Bottleneck)	**88** (16-186)	**38,790** (35,992-51,671)	**19,654** (10,538-24,016)	**24,383** (23,931-24,740)	**2,705** (816-7,011)	**-61,539.83**	**283,411**	**0**	**0.96**
Population 83									
(1) Standard (Stable)	-	-	18,731	-	-	-79,607.33	366,609	15,110	0.00
(2) Bottleneck + Expansion	43,249	379	37,336	43,193	24,322	-76,346.17	351,597	98	0.00
(3) Expansion + Bottleneck	378	47,039	37,234	24,509	23,827	-76,346.35	351,598	99	0.00
(4) No parameter restriction (Expansion)	**6** (5-186)	**18,908** (14,040-26,503)	**40,772** (39,541-43,491)	**25,426** (24,900-25,811)	**13,070** (8,262-16,201)	**-76,324.92**	**351,499**	**0**	**1.00**

Front									
Demographic model	**N_e ANC**	**N_e INT**	**N_e CUR**	**T ANC**	**T REC**	**$Log_{10}L$**	**AIC**	**ΔAIC**	**w_iAIC**
Population 01									
(1) Standard (Stable)	-	-	23,330	-	-	-45,074.22	207,578	8,056	0.00
(2) Bottleneck + Expansion	40,616	299	30,047	55,836	20,463	-43,324.25	199,526	4	0.13
(3) Expansion + Bottleneck	274	44,216	30,046	20,501	17,268	-43,324.27	199,526	4	0.12
(4) No parameter restriction (Expansion)	**29** (2-270)	**12,112** (4,109-35,358)	**31,251** (29,676-39,736)	**21,183** (20,327-21,491)	**15,677** (1,693-18,783)	**-43,323.48**	**199,522**	**0**	**0.75**
Population 27									
(1) Standard (Stable)	-	-	15,206	-	-	-77,158.01	355,330	17,021	0.00
(2) Bottleneck + Expansion	41,420	503	39,054	62,931	20,407	-73,495.57	338,470	161	0.00
(3) Expansion + Bottleneck	542	47,548	38,878	20,364	18,526	-73,496.36	338,473	165	0.00
(4) No parameter restriction (Expansion)	**68** (4-199)	**22,271** (17,077-25,785)	**47,718** (44,665-51,429)	**21,665** (21,162-22,066)	**7,558** (5,702-10,475)	**-73,460.59**	**338,309**	**0**	**1.00**
Population 67									
(1) Standard (Stable)	-	-	16,231	-	-	-73,895.09	340,303	11,653	0.00
(2) Bottleneck + Expansion	40,107	13	27,365	89,703	21,317	-71,365.92	328,662	12	0.00
(3) Expansion + Bottleneck	7	31,386	26,957	21,298	14,982	-71,366.08	328,663	13	0.00
(4) No parameter restriction (Expansion)	**5** (2-73)	**26,198** (21,342-26,688)	**47,053** (27,690-62,429)	**21,530** (21,014-21,673)	**413** (274-11,827)	**-71,363.27**	**328,650**	**0**	**1.00**

Population 69									
(1) Standard (Stable)	-	-	9,510	-	-	-58,732.42	270,477	7,054	0.00
(2) Bottleneck + Expansion	42,178	72	18,868	60,583	17,987	-57,221.05	263,523	100	0.00
(3) Expansion + Bottleneck	546	24,230	10,140	16,589	804	-57,200.42	263,428	5	0.07
(4) No parameter restriction (Expansion + Bottleneck)	**411** (304-697)	**22,089** (21,618-46,247)	**8,334** (5,661-16,835)	**17,091** (16,349-17,491)	**386** (190-6,834)	**-57,199.32**	**263,423**	**0**	**0.93**
Population 72									
(1) Standard (Stable)	-	-	47,530	-	-	-52,737.93	242,871	9,000	0.00
(2) Bottleneck + Expansion	19,825	46	25,194	46,431	18,816	-50,791.10	233,912	40	0.00
(3) Expansion + Bottleneck	88	28,705	18,726	18,762	1,693	-50,784.08	233,879	8	0.02
(4) No parameter restriction (Expansion + Bottleneck)	**72** (7-178)	**29,059** (27,667-37,830)	**5,738** (3,723-22,244)	**18,974** (18,458-19,237)	**186** (109-6,351)	**-50,782.35**	**233,871**	**0**	**0.98**
Population 74									
(1) Standard (Stable)	-	-	25,665	-	-	-57,850.41	266,415	9,867	0.00
(2) Bottleneck + Expansion	19,714	314	28,463	66,454	20,226	-55,726.03	256,638	90	0.00
(3) Expansion + Bottleneck	406	35,559	8,840	19,943	488	-55,707.60	256,553	5	0.08
(4) No parameter restriction (Expansion + Bottleneck)	**350** (232-490)	**34,394** (33,320-42,841)	**1,384** (1,133-21,568)	**20,225** (19,738-20,490)	**51** (41-3,628)	**-55,706.56**	**256,548**	**0**	**0.92**
Population 78									

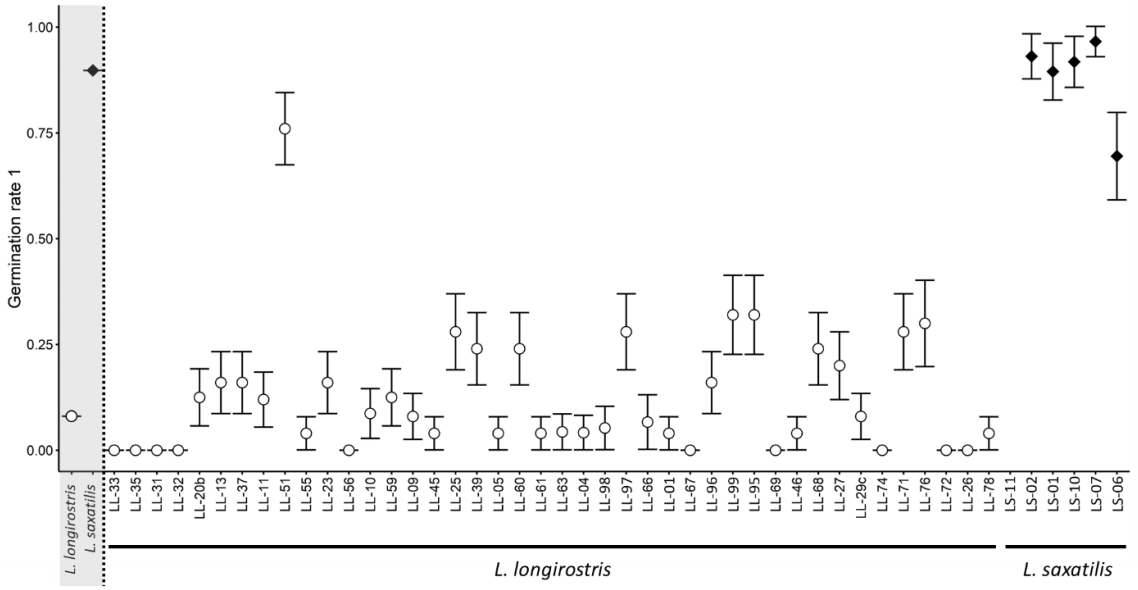
(1) Standard (Stable)	-	-	9,535	-	-	-56,332.59	259,425	10,393	0.00
(2) Bottleneck + Expansion	20,868	72	28,518	35,087	19,676	-54,075.46	249,037	4	0.10
(3) Expansion + Bottleneck	64	43,916	28,092	19,699	15,876	-54,075.87	249,039	6	0.04
(4) No parameter restriction (Expansion + Bottleneck)	**107** (29-220)	**29,809** (28,784-57,489)	**18,529** (12,449-27,911)	**19,734** (19,337-20,068)	**471** (304-13,664)	**-54,074.53**	**249,032**	**0**	**0.86**

— Supplementary Material —

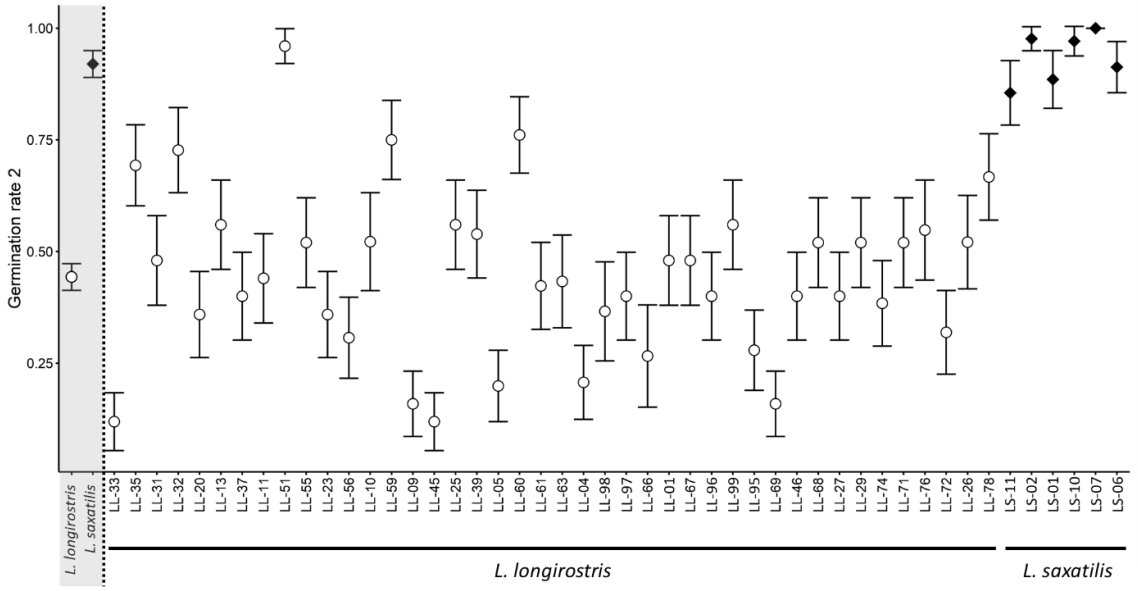
Chapter 2

Figure S2.1. Expected marginal mean trait values ± SE from linear mixed models for *Leontodon longirostris* and *L. saxatilis* populations grown in the common garden experiment. Populations for each species in the X -axis are sorted by latitude, from left (south) to right (north). Mean values for both species are also given (grey area). Population codes are the same that in Table S2.1 and S2.2. RCD: root crown diameter; RVB: reproductive to vegetative biomass ratio.

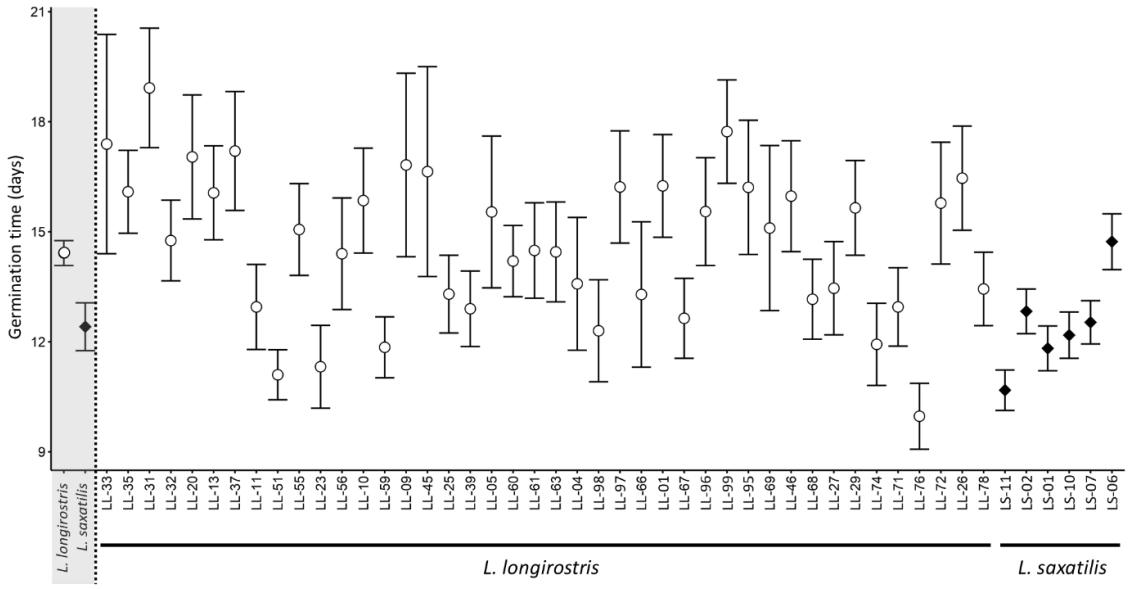
A)



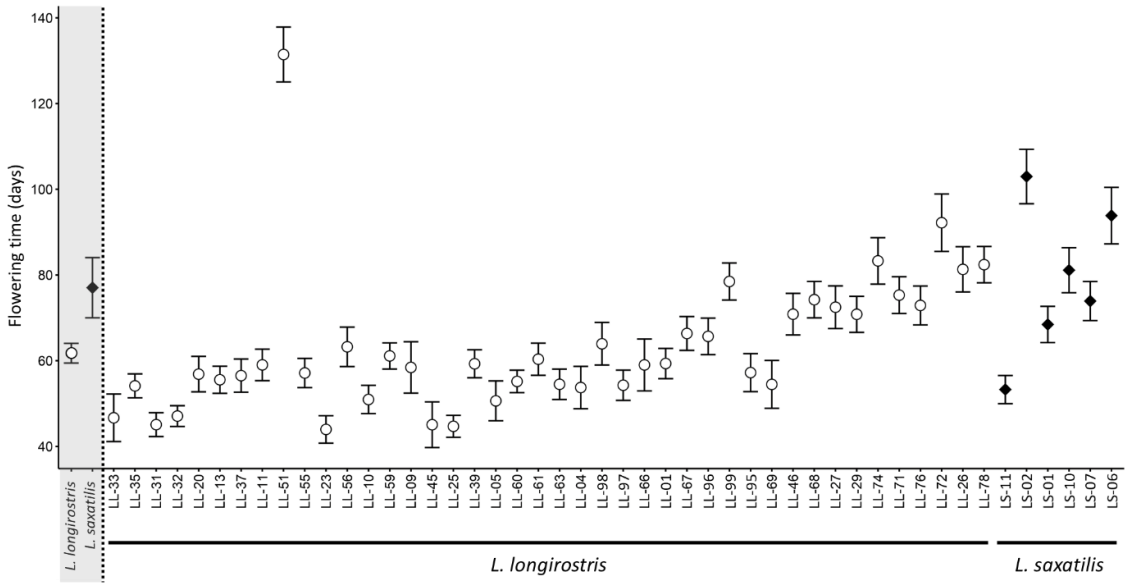
B)



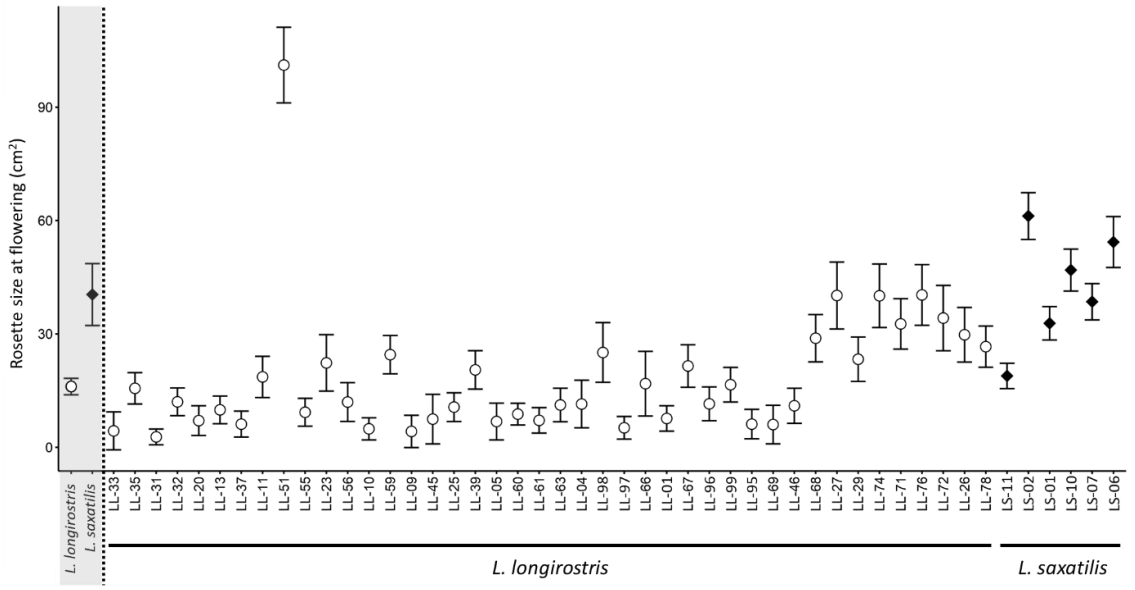
C)


Germination time (days)
21
18
15
12
9
L. longirostris
L. saxatilis
LL-33
LL-35
LL-31
LL-32
LL-20
LL-13
LL-37
LL-11
LL-51
LL-55
LL-23
LL-56
LL-10
LL-59
LL-09
LL-45
LL-25
LL-39
LL-05
LL-60
LL-61
LL-63
LL-04
LL-98
LL-97
LL-66
LL-01
LL-67
LL-96
LL-99
LL-95
LL-69
LL-46
LL-68
LL-27
LL-29
LL-74
LL-71
LL-76
LL-72
LL-26
LL-78
LS-11
LS-02
LS-01
LS-10
LS-07
LS-06
L. longirostris
L. saxatilis

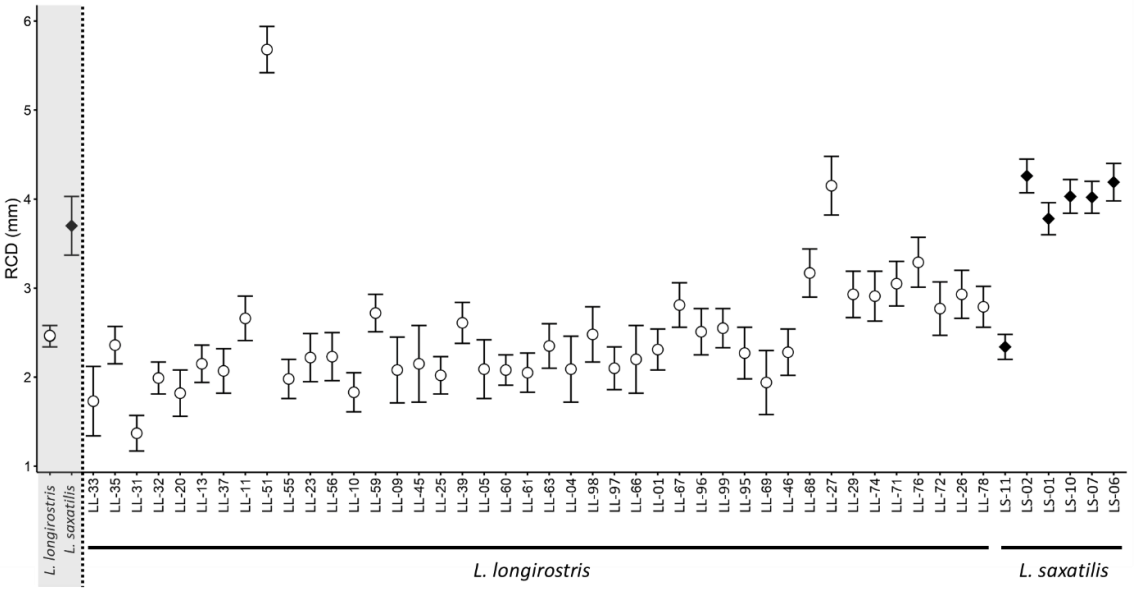
D)


Flowering time (days)
140
120
100
80
60
40
L. longirostris
L. saxatilis
LL-33
LL-35
LL-31
LL-32
LL-20
LL-13
LL-37
LL-11
LL-51
LL-55
LL-23
LL-56
LL-10
LL-59
LL-09
LL-45
LL-25
LL-39
LL-05
LL-60
LL-61
LL-63
LL-04
LL-98
LL-97
LL-66
LL-01
LL-67
LL-96
LL-99
LL-95
LL-69
LL-46
LL-68
LL-27
LL-29
LL-74
LL-71
LL-76
LL-72
LL-26
LL-78
LS-11
LS-02
LS-01
LS-10
LS-07
LS-06
L. longirostris
L. saxatilis

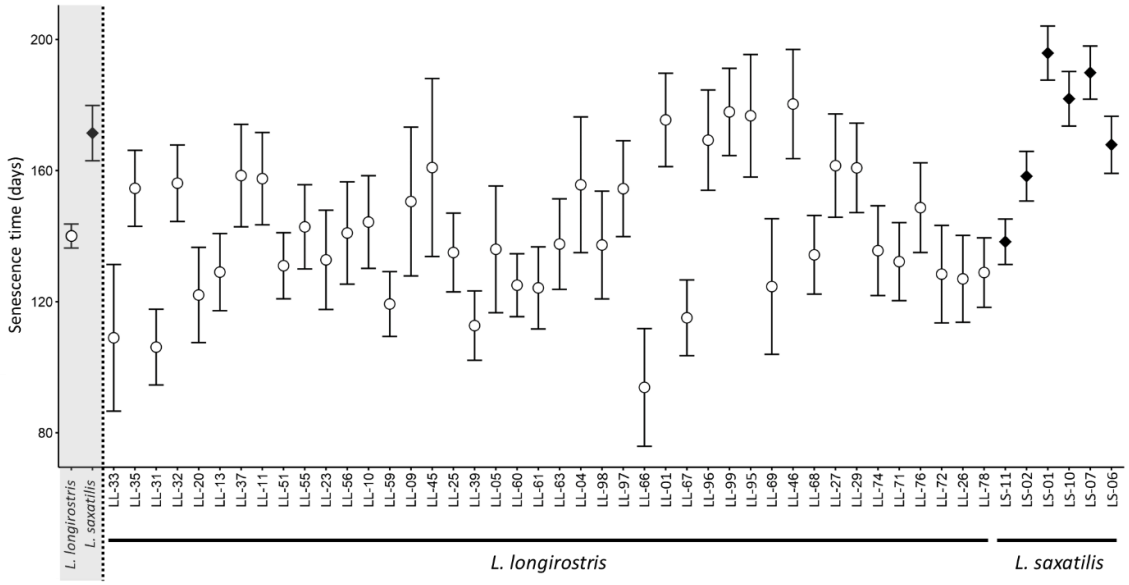
E)


Rosette size at flowering (cm²)
0
30
60
90
L. longirostris
L. saxatilis
LL-33
LL-35
LL-31
LL-32
LL-20
LL-13
LL-37
LL-11
LL-51
LL-55
LL-23
LL-56
LL-10
LL-59
LL-09
LL-45
LL-25
LL-39
LL-05
LL-60
LL-61
LL-63
LL-04
LL-98
LL-97
LL-66
LL-01
LL-67
LL-96
LL-99
LL-95
LL-69
LL-46
LL-68
LL-27
LL-29
LL-74
LL-71
LL-76
LL-72
LL-26
LL-78
LS-11
LS-02
LS-01
LS-10
LS-07
LS-06
L. longirostris
L. saxatilis

F)


RCD (mm)
1
2
3
4
5
6
L. longirostris
L. saxatilis
LL-33
LL-35
LL-31
LL-32
LL-20
LL-13
LL-37
LL-11
LL-51
LL-55
LL-23
LL-56
LL-10
LL-59
LL-09
LL-45
LL-25
LL-39
LL-05
LL-60
LL-61
LL-63
LL-04
LL-98
LL-97
LL-66
LL-01
LL-67
LL-96
LL-99
LL-95
LL-69
LL-46
LL-68
LL-27
LL-29
LL-74
LL-71
LL-76
LL-72
LL-26
LL-78
LS-11
LS-02
LS-01
LS-10
LS-07
LS-06
L. longirostris
L. saxatilis

G)


Senescence time (days)
200
160
120
80
L. longirostris
L. saxatilis
LL-33
LL-35
LL-31
LL-32
LL-20
LL-13
LL-37
LL-11
LL-51
LL-55
LL-23
LL-56
LL-10
LL-59
LL-09
LL-45
LL-25
LL-39
LL-05
LL-60
LL-61
LL-63
LL-04
LL-98
LL-97
LL-66
LL-01
LL-67
LL-96
LL-99
LL-95
LL-69
LL-46
LL-68
LL-27
LL-29
LL-74
LL-71
LL-76
LL-72
LL-26
LL-78
LS-11
LS-02
LS-01
LS-10
LS-07
LS-06
L. longirostris
L. saxatilis

H)

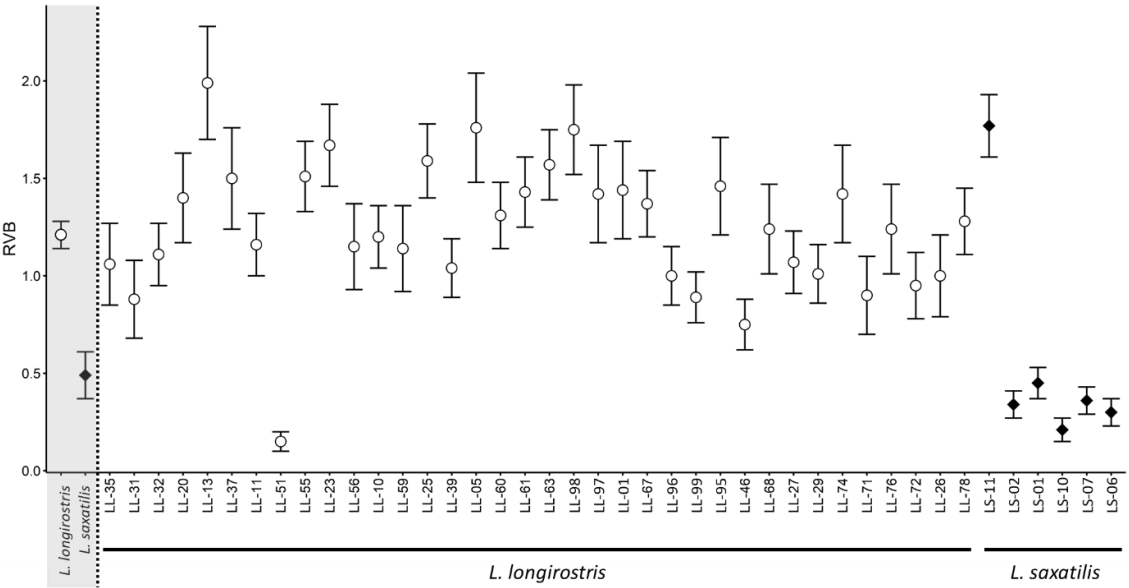

RVB
2.0
1.5
1.0
0.5
0.0
L. longirostris
L. saxatilis
LL-35
LL-31
LL-32
LL-20
LL-13
LL-37
LL-11
LL-51
LL-55
LL-23
LL-56
LL-10
LL-59
LL-25
LL-39
LL-05
LL-60
LL-61
LL-63
LL-98
LL-97
LL-01
LL-67
LL-96
LL-99
LL-95
LL-46
LL-68
LL-27
LL-29
LL-74
LL-71
LL-76
LL-72
LL-26
LL-78
LS-11
LS-02
LS-01
LS-10
LS-07
LS-06
L. longirostris
L. saxatilis

Figure S2.2. Correlations between flowering phenology scores in the field (2013 and 2014) and summer minimum temperature in *Leontodon longirostris*. Pearson's correlation coefficients (*r*) and associated *P*-values are given in each plot.

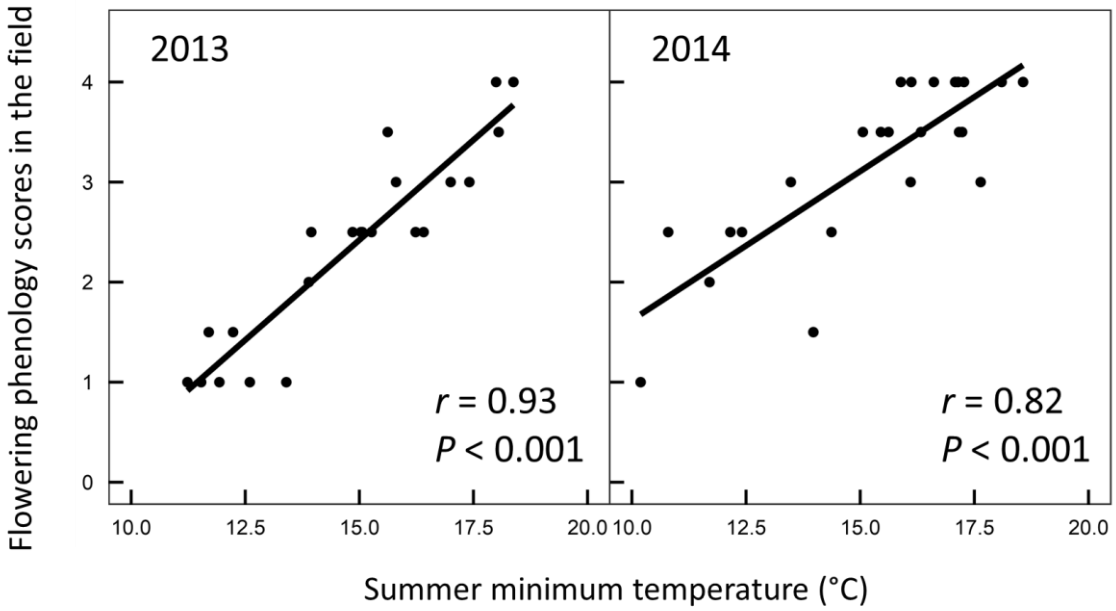


Table S2.1. List of geographic locations and identity of *Leontodon longirostris* and *L. saxatilis* used in this study. The type of data obtained in each location is also shown: FP = flowering phenology in the field; G1 = germination rate in the early fall experiment; G2 = germination rate in the late winter experiment; LT = Life-history traits; GD = genotypic data.

Population code	Population name (Province)	Latitude (º)	Longitude (º)	Altitude a.s.l. (m)	Type of data
L. longirostris					
LL-01	Mas de Barberans (Tarragona)	40.72477	0.39036	257	G1 / G2 / LT / GD
LL-95	Bellaterra (Barcelona)	41.50825	2.09743	174	FP / G1 / G2 / LT
LL-96	Fulleda (Lleida)	41.46885	1.02278	500	FP / G1 / G2 / LT
LL-97	La Galera (Tarragona)	40.65669	0.40793	176	G1 / G2 / LT
LL-98	Collado-Villalba (Madrid)	40.65001	-4.00717	890	G1 / G2 / LT
LL-02	Navacerrada (Madrid)	40.72026	-4.01665	1,185	FP
LL-03	San Ildefonso (Segovia)	40.90492	-4.03163	1,100	FP
LL-04	Majadahonda (Madrid)	40.46458	-3.86045	730	FP / G1 / G2 / LT / GD
LL-05	Almagro (Ciudad Real)	38.86955	-3.69605	650	FP / G1 / G2 / LT / GD
LL-06	Bolaños de Calatrava (Ciudad Real)	38.88183	-3.65220	680	FP
LL-07	Moral de Calatrava (Ciudad Real)	38.84151	-3.59971	800	FP
LL-09	Despeñaperros (Jaén)	38.39433	-3.50793	700	FP / G1 / G2 / LT
LL-10	Carboneros (Jaén)	38.22551	-3.63340	390	FP / G1 / G2 / LT / GD

LL-11	Arahal (Sevilla)	37.24962	-5.55369	90	FP / G1 / G2 / LT
LL-12	Antequera (Málaga)	37.07799	-4.47161	490	FP
LL-13	Cenes de la Vega (Granada)	37.15303	-3.53185	750	FP / G1 / G2 / LT
LL-15	Pinos Genil (Granada)	37.15255	-3.48393	1,120	FP
LL-16	Güéjar Sierra (Granada)	37.14011	-3.47164	1,390	FP
LL-17	Güéjar Sierra (Granada)	37.13894	-3.45868	1,500	FP
LL-18	Güéjar Sierra (Granada)	37.13652	-3.47380	1,500	FP
LL-18b	Güéjar Sierra (Granada)	37.13606	-3.47576	1,470	FP
LL-19	Monachil (Granada)	37.13767	-3.49926	1,360	FP
LL-20	Monachil (Granada)	37.13782	-3.50930	1,200	FP / G2 / LT / GD
LL-20b	Monachil (Granada)	37.13666	-3.51353	1,150	FP / G1
LL-21	Monachil (Granada)	37.13222	-3.52650	930	FP
LL-22	Baza (Granada)	37.44324	-2.89237	1,170	FP
LL-23	La Hoya (Murcia)	37.71190	-1.60044	300	FP / G1 / G2 / LT / GD
LL-25	Dos Torres (Córdoba)	38.46917	-4.93361	550	G1 / G2 / LT / GD
LL-26	La Jonquera (Girona)	42.39860	2.90503	180	G1 / G2 / LT
LL-27	Tagamanent (Barcelona)	41.77577	2.32843	1,200	G1 / G2 / LT / GD
LL-29	Seva (Barcelona)	41.82759	2.28836	760	G2 / LT
LL-29c	Seva (Barcelona)	41.82561	2.28785	774	G1
LL-31	Salobreña (Granada)	36.74531	-3.60505	42	G1 / G2 / LT / GD
LL-32	Vélez-Málaga (Málaga)	36.75302	-4.10286	23	G1 / G2 / LT

LL-33	Estepona (Málaga)	36.46500	-5.11729	87	G1 / G2 / LT
LL-35	Medina-Sidonia (Cádiz)	36.46741	-5.94530	104	FP / G1 / G2 / LT /GD
LL-37	Peñuelas (Granada)	37.19238	-3.85408	550	FP / G1 / G2 / LT
LL-39	Muro (Alicante)	38.78147	-0.44664	419	FP / G1 / G2 / LT
LL-45	El Campello (Alicante)	38.43531	-0.39196	18	G1 / G2 / LT / GD
LL-46	Sant Celoni (Barcelona)	41.66983	2.46166	171	G1 / G2 / LT
LL-51	Baza (Granada)	37.50851	-2.76637	801	FP / G1 / G2 / LT
LL-52	Cenes de la Vega (Granada)	37.14470	-3.48758	1,275	FP
LL-53	Almárgen (Málaga)	37.01718	-4.94450	387	FP
LL-54	Zahara de la Sierra (Cádiz)	36.84632	-5.39894	360	FP
LL-55	Cantillana (Sevilla)	37.60393	-5.85401	20	FP / G1 / G2 / LT /GD
LL-56	Alanís (Sevilla)	38.01451	-5.74034	580	FP / G1 / G2 / LT
LL-57	Argallón (Córdoba)	38.20069	-5.48087	631	FP
LL-58	Berlanga (Badajoz)	38.27238	-5.81480	553	FP
LL-59	Villagarcía de la Torre (Badajoz)	38.29318	-6.08706	597	FP / G1 / G2 / LT
LL-60	Mérida (Cáceres)	38.90572	-6.36792	256	FP / G1 / G2 / LT /GD
LL-61	Cáceres (Cáceres)	39.42090	-6.35615	476	FP / G1 / G2 / LT
LL-62	Cañaveral (Cáceres)	39.77155	-6.42112	250	FP
LL-63	Plasencia (Cáceres)	40.08421	-6.07815	391	FP / G1 / G2 / LT / GD
LL-64	Baños de Montemayor (Cáceres)	40.32349	-5.86168	811	FP
LL-65	Guijuelo (Salamanca)	40.52126	-5.67120	950	FP

LL-66	Guijuelo (Salamanca)	40.66950	-5.61765	887	FP / G1 / G2 / LT
LL-67	Alaejos (Valladolid)	41.28169	-5.24706	778	FP / G1 / G2 / LT / GD
LL-68	Calatañazor (Soria)	41.70568	-2.78855	1,106	FP / G1 / G2 / LT / GD
LL-69	Malgrat de Mar (Barcelona)	41.66086	2.74070	20	G1 / G2 / LT / GD
LL-71	Hontoria del Pinar (Burgos)	41.85812	-3.15584	1,063	G1 / G2 / LT
LL-72	Hontoria de la Cantera (Burgos)	42.18096	-3.63403	964	G1 / G2 / LT / GD
LL-74	Santa Eulalia de Tábara (Zamora)	41.85155	-5.79235	717	G1 / G2 / LT / GD
LL-76	Espadañeda (Zamora)	42.09969	-6.41207	1,029	G1 / G2 / LT
LL-78	La virgen del Camino (León)	42.57470	-5.62874	897	G1 / G2 / LT / GD
LL-99	Bellaterra (Barcelona)	41.50006	2.10330	151	G1 / G2 / LT
L. saxatilis					
LS-01	La Jonquera (Girona)	42.39695	2.90859	170	G1 / G2 / LT
LS-02	Gallur (Zaragoza)	41.83458	-1.32116	276	G1 / G2 / LT
LS-06	Colunga (Asturias)	43.49981	-5.26323	13	G1 / G2 / LT
LS-07	Noja (Cantabria)	43.47767	-3.51605	1	G1 / G2 / LT
LS-10	Arenas de Iguña (Cantabria)	43.18314	-4.05902	216	G1 / G2 / LT
LS-11	Mazagón (Huelva)	37.13250	-6.67083	67	G2 / LT

Table S2.2. Mean values (± SD) for phenotypic traits measured on *Leontodon longirostris* and *L. saxatilis* populations. Germination rates 1 and 2 correspond to early fall and late winter experiments, respectively. For each trait, the total number of populations, total number of families, and range of families per population are indicated. RCD: root crown diameter; RVB: reproductive to vegetative biomass ratio.

Population code	Germination rate 1	Germination rate 2	Germination time (days)	Flowering time (days)	Rosette size at flowering (cm^2)	RCD (mm)	Senescence time (days)	RVB
L. longirostris								
Total no. populations Total no. families (No. of families per population)	42 1016 (15-25)	42 1020 (15-26)	42 457 (3-24)	42 435 (3-19)	42 434 (3-19)	42 435 (3-24)	42 435 (3-19)	36 273 (5-11)
LL-01	0.04 ± 0.20	0.48 ± 0.25	16.67 ± 3.42	60.67 ± 9.77	10.66 ± 13.58	2.38 ± 0.72	178.83 ± 43.82	1.58 ± 1.20
LL-95	0.32 ± 0.47	0.28 ± 0.20	16.71 ± 4.15	58.57 ± 12.78	6.79 ± 6.66	2.22 ± 0.53	185.00 ± 45.59	1.45 ± 0.21
LL-96	0.16 ± 0.37	0.40 ± 0.24	15.90 ± 3.87	66.10 ± 13.57	16.36 ± 19.73	2.61 ± 1.08	172.90 ± 46.89	1.01 ± 0.31
LL-97	0.28 ± 0.45	0.40 ± 0.24	16.70 ± 4.45	55.20 ± 9.76	7.40 ± 9.11	2.16 ± 0.79	157.10 ± 49.58	1.44 ± 0.48
LL-98	0.05 ± 0.22	0.37 ± 0.23	12.57 ± 3.46	67.57 ± 20.78	32.23 ± 37.06	2.44 ± 0.67	141.43 ± 49.13	1.81 ± 0.69
LL-04	0.04 ± 0.20	0.21 ± 0.16	15.00 ± 7.25	54.00 ± 7.00	14.21 ± 15.82	2.1 ± 1.21	163.20 ± 38.89	-
LL-05	0.04 ± 0.20	0.20 ± 0.16	16.20 ± 4.97	52.80 ± 14.36	8.38 ± 10.21	2.08 ± 0.41	141.60 ± 31.99	1.82 ± 0.80
LL-09	0.08 ± 0.27	0.16 ± 0.13	16.75 ± 3.30	59.00 ± 7.26	6.00 ± 5.74	2.13 ± 0.48	157.50 ± 45.14	-
LL-10	0.09 ± 0.28	0.52 ± 0.25	16.36 ± 4.78	51.80 ± 12.53	8.54 ± 13.00	1.88 ± 0.64	146.20 ± 35.91	1.26 ± 0.48
LL-11	0.12 ± 0.32	0.44 ± 0.25	13.55 ± 4.59	60.91 ± 16.07	28.38 ± 33.79	2.95 ± 1.42	162.82 ± 50.67	1.23 ± 0.66

LL-13	0.16 ± 0.37	0.56 ± 0.25	17.00 ± 6.37	55.62 ± 8.38	13.30 ± 15.49	2.23 ± 0.86	131.46 ± 38.88	2.04 ± 0.79
LL-20	-	0.36 ± 0.23	17.56 ± 3.61	57.00 ± 10.24	9.75 ± 11.96	1.91 ± 0.94	127.38 ± 48.23	1.43 ± 0.51
LL-20b	0.13 ± 0.33	-	-	-	-	-	-	-
LL-23	0.16 ± 0.37	0.36 ± 0.23	12.22 ± 5.61	44.00 ± 2.93	28.38 ± 21.55	2.37 ± 0.95	134.25 ± 40.37	1.72 ± 0.63
LL-25	0.28 ± 0.45	0.56 ± 0.25	14.21 ± 6.19	45.08 ± 7.44	16.04 ± 17.93	2.17 ± 0.69	136.54 ± 43.53	1.62 ± 0.43
LL-26	0.00 ± 0.00	0.52 ± 0.25	17.17 ± 4.86	82.30 ± 11.20	34.27 ± 26.50	2.94 ± 1.07	130.10 ± 49.41	1.02 ± 0.43
LL-27	0.20 ± 0.40	0.40 ± 0.24	13.70 ± 3.09	75.89 ± 29.98	49.08 ± 39.38	4.45 ± 1.73	161.22 ± 41.33	1.13 ± 0.48
LL-29	-	0.52 ± 0.25	16.08 ± 3.52	73.75 ± 17.99	25.82 ± 17.82	2.89 ± 0.71	164.42 ± 38.53	1.02 ± 0.21
LL-29c	0.08 ± 0.27	-	-	-	-	-	-	-
LL-31	0.00 ± 0.00	0.48 ± 0.25	19.42 ± 4.36	46.00 ± 5.93	2.61 ± 2.58	1.32 ± 0.29	115.27 ± 64.02	0.98 ± 0.56
LL-32	0.00 ± 0.00	0.73 ± 0.20	15.88 ± 6.48	47.25 ± 8.02	16.69 ± 18.29	2.02 ± 0.51	163.88 ± 59.89	1.19 ± 0.63
LL-33	0.00 ± 0.00	0.12 ± 0.11	17.33 ± 2.08	47.00 ± 1.73	4.32 ± 6.09	1.58 ± 0.50	110.33 ± 29.30	-
LL-35	0.00 ± 0.00	0.69 ± 0.21	16.94 ± 5.14	55.19 ± 10.89	22.25 ± 24.83	2.50 ± 0.97	155.88 ± 47.55	1.06 ± 0.35
LL-37	0.16 ± 0.37	0.40 ± 0.24	18.50 ± 7.68	57.33 ± 12.44	7.26 ± 6.27	2.03 ± 0.48	161.11 ± 48.47	1.50 ± 0.44
LL-39	0.24 ± 0.43	0.54 ± 0.25	13.57 ± 4.38	60.50 ± 10.96	24.46 ± 20.72	2.54 ± 0.76	117.57 ± 46.74	1.10 ± 0.56
LL-45	0.04 ± 0.20	0.12 ± 0.11	17.33 ± 4.62	45.67 ± 3.51	9.52 ± 12.36	2.12 ± 0.50	170.67 ± 74.57	-
LL-46	0.04 ± 0.20	0.40 ± 0.24	16.50 ± 4.03	72.00 ± 12.44	14.32 ± 15.22	2.31 ± 0.41	182.33 ± 26.21	0.78 ± 0.29
LL-51	0.76 ± 0.43	0.96 ± 0.04	11.29 ± 2.07	135.22 ± 31.83	104.23 ± 28.17	5.67 ± 0.83	133.72 ± 32.33	0.15 ± 0.07
LL-55	0.04 ± 0.20	0.52 ± 0.25	15.38 ± 3.18	57.83 ± 9.85	14.79 ± 26.41	2.08 ± 0.71	143.08 ± 48.92	1.61 ± 0.94
LL-56	0.00 ± 0.00	0.31 ± 0.21	15.13 ± 5.57	64.88 ± 11.72	15.17 ± 20.08	2.23 ± 0.78	147.38 ± 45.39	1.18 ± 0.43

LL-59	0.13 ± 0.33	0.75 ± 0.19	12.28 ± 3.20	62.53 ± 13.27	28.29 ± 18.88	2.86 ± 1.15	123.82 ± 42.29	1.19 ± 0.42
LL-60	0.24 ± 0.43	0.76 ± 0.18	14.95 ± 4.44	56.00 ± 9.04	13.6 ± 16.39	2.18 ± 0.96	127.53 ± 32.70	1.36 ± 0.49
LL-61	0.04 ± 0.20	0.42 ± 0.24	15.09 ± 4.32	61.27 ± 10.71	11.42 ± 17.6	2.16 ± 1.00	123.73 ± 40.12	1.49 ± 0.60
LL-63	0.04 ± 0.20	0.43 ± 0.25	15.30 ± 5.50	55.70 ± 9.27	13.88 ± 16.21	2.33 ± 0.61	143.30 ± 48.57	1.60 ± 0.42
LL-66	0.07 ± 0.25	0.27 ± 0.20	13.50 ± 3.11	58.50 ± 6.95	18.80 ± 14.44	2.30 ± 0.77	99.25 ± 35.72	-
LL-67	0.00 ± 0.00	0.48 ± 0.25	13.08 ± 3.68	67.58 ± 16.71	26.02 ± 20.02	2.88 ± 1.02	119.00 ± 33.66	1.45 ± 0.79
LL-68	0.24 ± 0.43	0.52 ± 0.25	14.00 ± 5.26	76.38 ± 19.05	33.04 ± 21.62	3.19 ± 1.01	139.23 ± 42.73	1.27 ± 0.48
LL-69	0.00 ± 0.00	0.16 ± 0.13	15.25 ± 3.10	55.50 ± 11.70	7.53 ± 7.27	2.01 ± 0.27	129.00 ± 52.52	-
LL-71	0.28 ± 0.45	0.52 ± 0.25	13.38 ± 3.43	77.23 ± 15.62	39.4 ± 29.26	3.23 ± 1.25	133.92 ± 42.75	0.92 ± 0.37
LL-72	0.00 ± 0.00	0.32 ± 0.22	15.88 ± 3.18	98.25 ± 45.89	43.18 ± 40.54	2.93 ± 1.21	133.13 ± 50.02	0.96 ± 0.25
LL-74	0.00 ± 0.00	0.38 ± 0.24	12.50 ± 4.33	84.80 ± 20.91	45.52 ± 26.43	2.98 ± 0.87	139.90 ± 25.68	1.47 ± 0.62
LL-76	0.30 ± 0.46	0.55 ± 0.25	10.27 ± 3.07	74.18 ± 21.51	49.06 ± 34.35	3.46 ± 1.13	154.18 ± 51.43	1.33 ± 0.64
LL-78	0.04 ± 0.20	0.67 ± 0.22	13.81 ± 3.15	84.63 ± 17.17	28.72 ± 17.46	2.87 ± 0.75	131.88 ± 43.65	1.33 ± 0.54
LL-99	0.32 ± 0.47	0.56 ± 0.25	18.07 ± 3.17	79.71 ± 16.47	19.84 ± 14.33	2.61 ± 0.48	176.79 ± 26.45	0.90 ± 0.20

L. saxatilis

Total no. populations Total no. families (No. of families per population)	5 122 (22-25)	6 148 (22-26)	6 134 (21-25)	6 119 (16-21)	6 117 (15-21)	6 119 (17-21)	6 118 (16-21)	6 30 (5)
LS-01	0.90 ± 0.32	0.89 ± 0.13	12.10 ± 2.86	71.10 ± 22.95	33.63 ± 10.45	3.80 ± 0.54	196.43 ± 21.97	0.46 ± 0.13

LS-02	0.93 ± 0.27	0.98 ± 0.04	13.12 ± 2.74	107.90 ± 35.78	66.53 ± 42.36	4.34 ± 1.22	159.80 ± 36.04	0.36 ± 0.16
LS-06	0.70 ± 0.47	0.91 ± 0.11	15.19 ± 3.59	96.50 ± 25.61	56.51 ± 21.86	4.24 ± 1.14	169.50 ± 25.44	0.31 ± 0.13
LS-07	0.97 ± 0.20	1.00 ± 0.00	12.68 ± 1.95	76.52 ± 23.74	40.89 ± 20.44	4.05 ± 0.68	191.05 ± 23.20	0.38 ± 0.17
LS-10	0.92 ± 0.29	0.97 ± 0.04	12.52 ± 2.99	85.11 ± 27.46	49.98 ± 30.95	4.04 ± 0.48	183.16 ± 28.98	0.23 ± 0.12
LS-11	-	0.86 ± 0.16	11.24 ± 4.16	53.29 ± 4.38	22.85 ± 16.34	2.40 ± 0.77	143.05 ± 53.03	1.79 ± 0.50

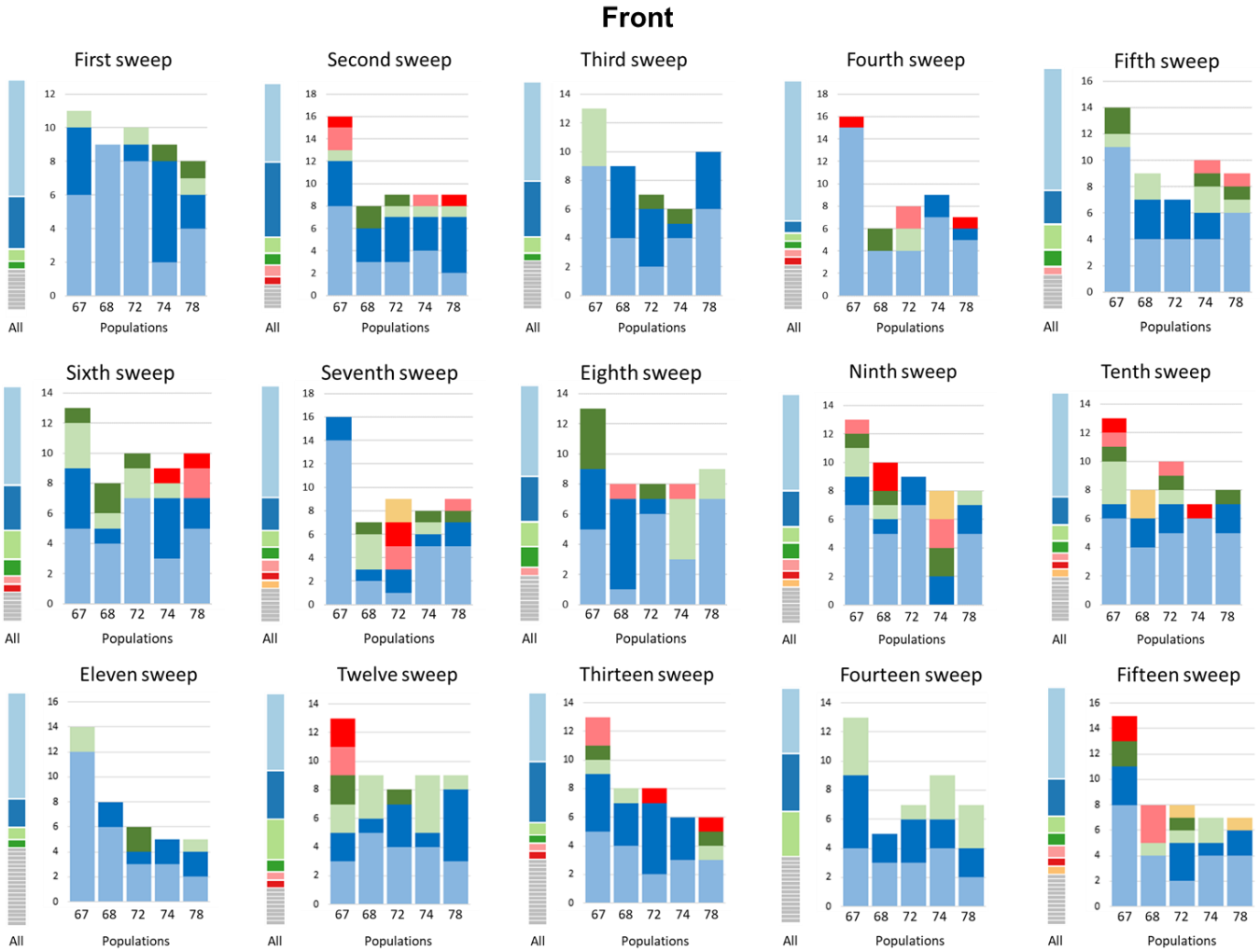
Table S2.3. Comparison of candidate models explaining the observed variability in phenotypic traits according to climatic conditions or/and genetic structure (ancestry) in *Leontodon longirostris*. Three models were compared for each trait using different predictors: "Climate", "Ancestry" and "Climate + Ancestry" (see Material and methods section). The best model for each trait was chosen based on the lowest AIC score (bold values). Climate refers the best correlated climatic variable and ancestry to the first two genetic PCA axes. A total of 20 populations with both phenotypic and genetic data were included in all traits except for RVD, for which information was available on 17 populations. Population LL-51, with extreme values, was excluded from the analyses. RCD: root crown diameter; RVB: reproductive to vegetative biomass ratio

	Climatic variable	**Climate**	**Ancestry**	**Climate + Ancestry**
Germination time	Summer min. temp.	**82.24**	82.76	83.44
Flowering time	Summer min. temp.	**137.47**	153.08	139.11
Rosette size at flowering	Summer min. temp.	**139.78**	154.14	142.31
RCD	Summer min. temp.	**22.20**	34.81	25.67
Senescence time	Summer prec.	173.83	176.69	**168.83**
RVB	Summer max. temp.	**-1.47**	0.73	0.04

— Supplementary Material —

Chapter 3

Figure S3.1. Multi-locus genotype (MLG) frequency spectra for the top 15 exclusive sweeps detected in each population of the *core* and *front* ranges with the G12 method. Unique MLGs in each population are not shown.



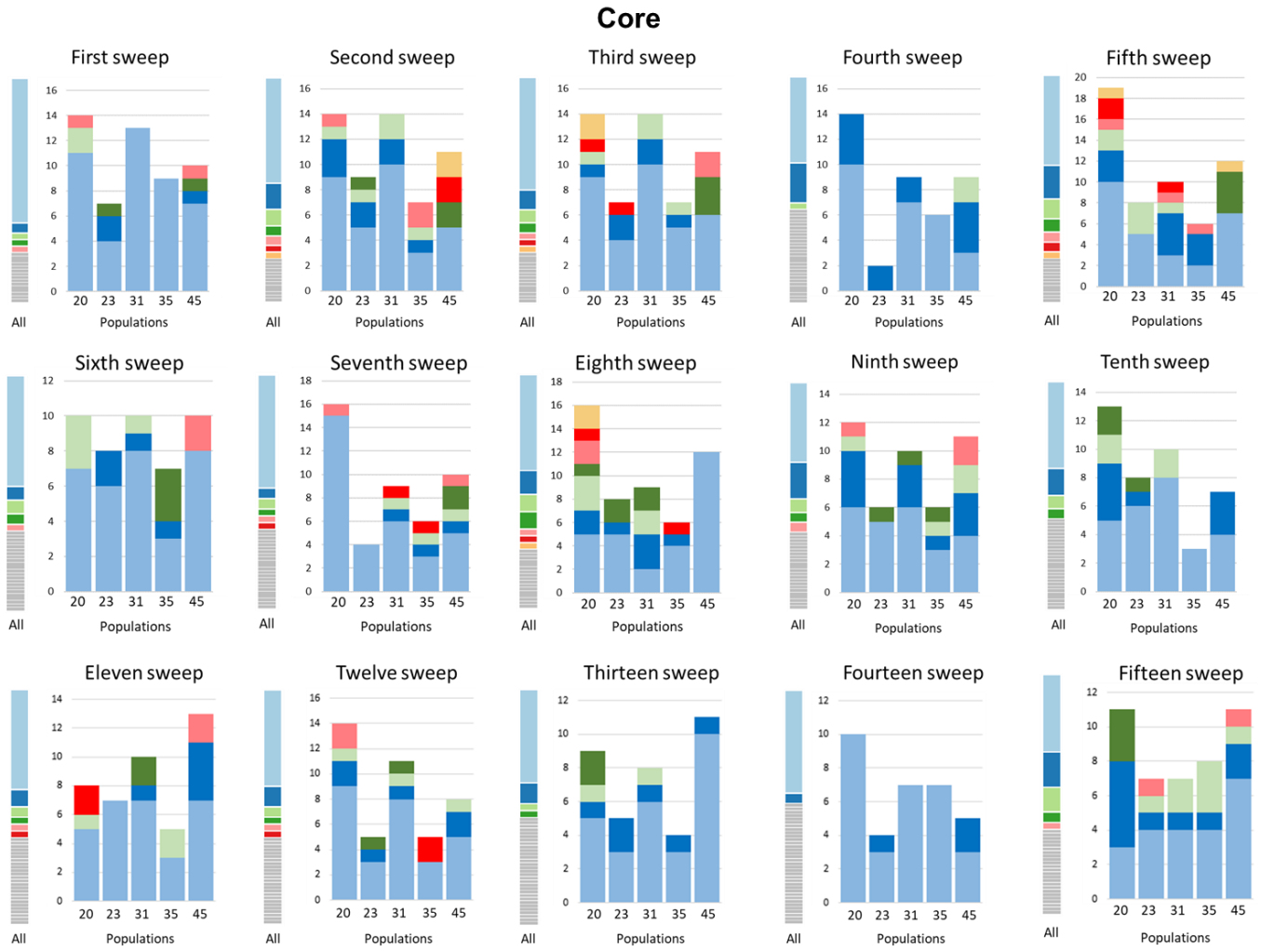

Core
First sweep
Second sweep
Third sweep
Fourth sweep
Fifth sweep
Sixth sweep
Seventh sweep
Eighth sweep
Ninth sweep
Tenth sweep
Eleven sweep
Twelve sweep
Thirteen sweep
Fourteen sweep
Fifteen sweep
All
20 23 31 35 45
Populations

Figure S3.2. Multi-locus genotype (MLG) frequency spectra for exclusive sweeps detected in *core* and *front* ranges with the G12 method. The height of the top light blue region in each bar indicates the frequency of the most prevalent MLG in the sample (68 and 57 individuals in *core* and *front*, respectively). Heights of subsequent colored bars indicate the frequency of the second, third, and so on, most-frequent MLG in the sample. Gray bars indicate unique MLGs. G12 is the expected MLG homozygosity combining the frequencies of the first and the second most common MLGs, and G2/G1 is the ratio between expected MLG including (G1) and excluding (G2) the most frequent MLG (see text for further details).

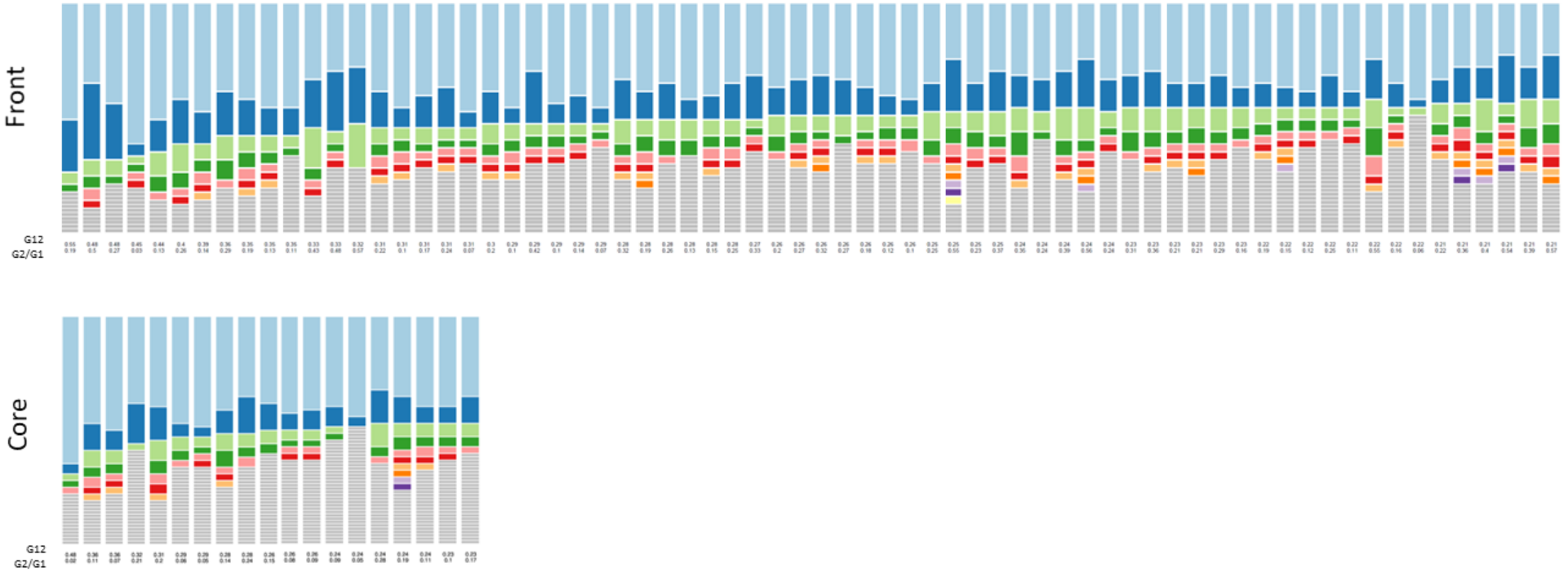


Figure S3.3. Multi-locus genotype (MLG) frequency spectra for shared sweeps between *core* and *front* ranges detected with the G12 method. The height of the top light blue region in each bar indicates the frequency of the most prevalent MLG in the sample (68 and 57 individuals in *core* and *front*, respectively). Heights of subsequent colored bars indicate the frequency of the second, third, and so on, most-frequent MLG in the sample. Gray bars indicate unique MLGs. G12 is the expected MLG homozygosity combining the frequencies of the first and the second most common MLGs, and G2/G1 is the ratio between expected MLG including (G1) and excluding (G2) the most frequent MLG (see text for further details).

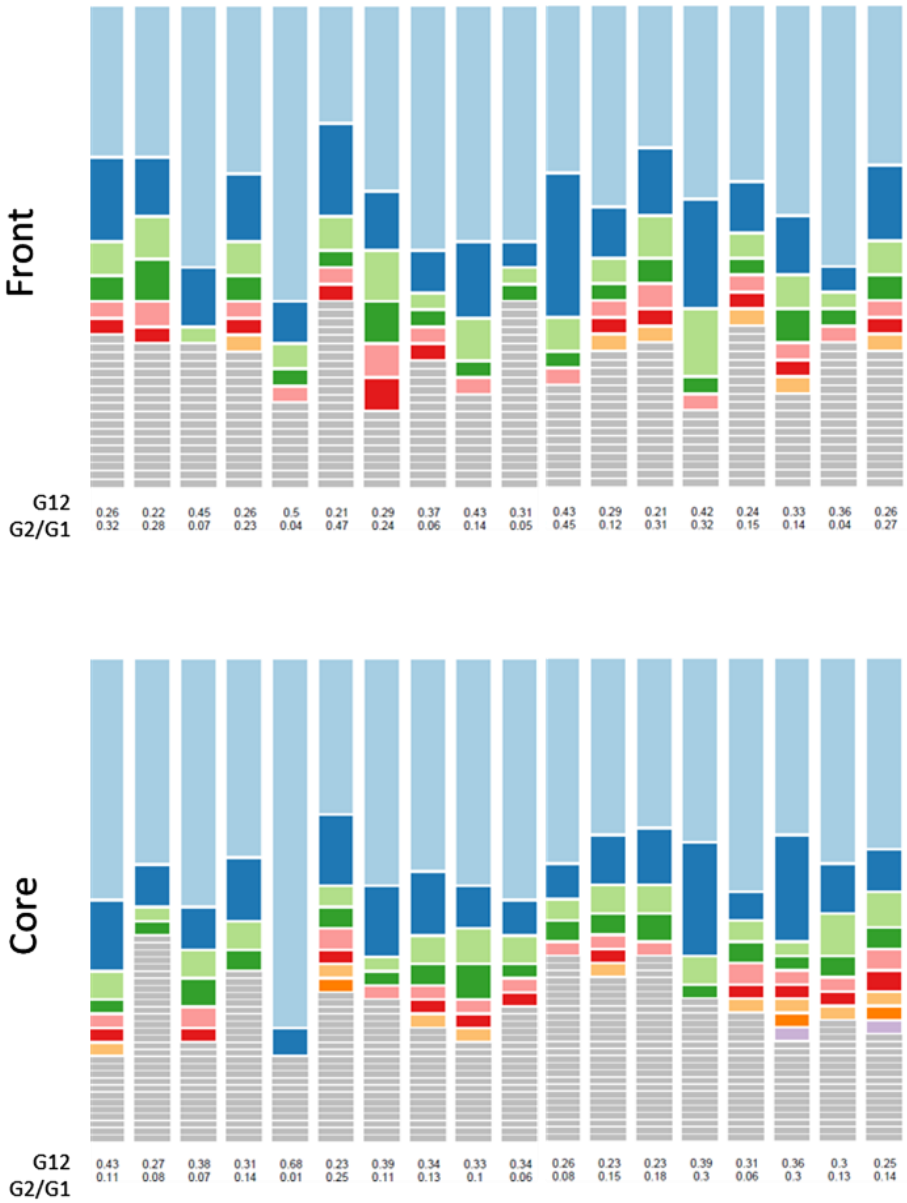


Table S3.1. Location, sample size, climatic and phenotypic information of the 21 *Leontodon longirostris* populations included in this study. Precipitation is the total accumulated amount during the period. *Data from de Pedro et al. (unpublished). *N*=sample size; *Nf*=number of samples corresponding to some of the maternal families in the common garden experiments.

						Climatic variables								**Phenotypic variables***			
Population code	***N***	***Nf***	**Latitude (º)**	**Longitude (º)**	**Altitude (m)**	**Min. Temp. Winter (ºC)**	**Min. Temp. Spring (ºC)**	**Max. Temp. Spring (ºC)**	**CV Prec. Spring**	**Prec. Winter (mm)**	**Prec. Spring (mm)**	**Prec. Summer (mm)**	**Prec. Autumn (mm)**	**Germination rate 1**	**Germination rate 2**	**Germination time (days)**	**Flowering time (days)**
01	6	-	40.72	0.39	257	3.9	8.5	19.9	0.84	130	161	93	200	0.04	0.48	16.67	60.67
04	9	-	40.46	-3.86	730	1.1	6.0	18.5	0.59	157	145	60	163	0.04	0.21	15.00	54.00
05	9	5	38.87	-3.70	650	1.2	6.6	19.5	0.65	139	129	41	122	0.04	0.20	16.20	52.80
10	10	4	38.23	-3.63	390	3.8	8.8	21.7	0.73	203	150	30	153	0.09	0.52	16.36	51.80
20	20	-	37.14	-3.51	1,200	1.7	6.0	17.9	0.69	204	155	29	157	0.13	0.36	17.56	57.00
23	10	4	37.71	-1.60	300	5.4	9.7	21.7	0.84	88	93	28	109	0.16	0.36	12.22	44.00
25	7	3	38.47	-4.93	550	3.2	7.9	20.5	0.68	174	142	38	147	0.28	0.56	14.21	45.08
27	14	2	41.78	2.33	1,200	-2.3	2.0	12.0	0.59	179	224	175	242	0.20	0.40	13.70	75.89
31	14	-	36.75	-3.61	54	10.4	13.2	19.8	0.95	208	117	10	149	0.00	0.48	19.42	46.00
35	10	10	36.47	-5.95	104	7.4	11.1	22.1	0.76	323	172	18	215	0.00	0.69	16.94	55.19
45	14	-	38.44	-0.39	18	7.7	11.3	20.5	0.83	101	95	32	140	0.04	0.12	17.33	45.67
55	10	8	37.60	-5.85	20	5.5	9.9	23.3	0.74	235	145	20	176	0.04	0.52	15.38	57.83
60	9	3	38.91	-6.37	256	4.2	8.9	21.9	0.71	187	137	32	158	0.24	0.76	14.95	56.00
63	9	5	40.08	-6.08	391	3.3	8.2	20.1	0.60	275	203	51	235	0.04	0.43	15.30	55.70
67	17	9	41.28	-5.25	778	0.1	4.5	17.0	0.58	132	121	57	131	0.00	0.48	13.08	67.58
68	10	4	41.71	-2.79	1,106	-1.8	2.3	14.8	0.51	177	174	105	170	0.24	0.52	14.00	76.38
69	17	2	41.66	2.74	20	4.3	8.9	19.3	0.71	148	173	122	231	0.00	0.16	15.25	55.50
72	10	3	42.18	-3.63	964	-0.9	3.1	14.7	0.48	190	184	104	183	0.00	0.32	15.88	98.25
74	10	6	41.85	-5.79	717	0.2	4.6	17.2	0.62	175	146	67	161	0.00	0.38	12.50	84.80
78	10	10	42.57	-5.63	897	-0.5	3.5	15.4	0.61	191	156	83	171	0.04	0.67	13.81	84.63
83	13	-	38.60	-0.49	1,021	1.9	5.5	16.4	0.69	123	127	55	146	-	-	-	-

Table S3.2. Selective sweeps detected in *core* and *front* populations with SweeD. Non-annotated sweeps correspond to upstream or downstream intergenic regions. Sweeps that were also detected with the G12 method are highlighted in grey. CLR = Composite likelihood ratio.

Group	Scaffold	Gene code	Gene start	Gene end	CLR position	CLR likelihood	Genes in sweep region (BLASTX annotation)
Exclusive							
Core	scaffold264195	maker-scaffold264195-exonerate_protein2genome-gene-0.20	6,856	9,960	9,763	13.57	probable alkaline/neutral invertase D
Core	scaffold387913	maker-scaffold387913-exonerate_protein2genome-gene-0.18	10,567	12,606	11,433	12.12	ABC transporter G family member 1-like
Core	scaffold377567	maker-scaffold377567-exonerate_protein2genome-gene-0.0	10,499	17,745	13,801	11.43	probable ubiquitin-conjugating enzyme E2 23 isoform X1
Core	scaffold377567	maker-scaffold377567-exonerate_protein2genome-gene-0.0	10,499	17,745	14,729	10.28	probable ubiquitin-conjugating enzyme E2 23 isoform X1
Core	scaffold377567	maker-scaffold377567-exonerate_protein2genome-gene-0.0	10,499	17,745	14,265	9.71	probable ubiquitin-conjugating enzyme E2 23 isoform X1
Core	scaffold264195	maker-scaffold264195-exonerate_protein2genome-gene-0.20	6,856	9,960	7,285	9.53	probable alkaline/neutral invertase D
Core	scaffold339125	maker-scaffold339125-exonerate_protein2genome-gene-0.0	4,228	9,162	4,823	8.85	chaperonin CPN60-2, mitochondrial-like
Core	scaffold300029	maker-scaffold300029-exonerate_protein2genome-gene-0.22	8,431	15,161	9,253	8.23	phosphatidylinositol 4-kinase alpha 1-like isoform X2
Core	scaffold362323				14,161	7.91	
Core	scaffold264195	maker-scaffold264195-exonerate_protein2genome-gene-0.20	6,856	9,960	9,417	7.71	probable alkaline/neutral invertase D
Core	scaffold269378				2,148	7.56	
Core	scaffold264195				10,574	7.49	

Core	scaffold302647	maker-scaffold302647-exonerate_protein2genome-gene-0.20	5,767	16,099	6,084	7.31	acyl-CoA-binding domain-containing protein 4
Core	scaffold264195				12,196	7.28	
Core	scaffold390292	maker-scaffold390292-exonerate_protein2genome-gene-0.58	15,210	19,017	17,881	7.16	autophagy-related protein 18g-like
Core	scaffold300029	maker-scaffold300029-exonerate_protein2genome-gene-0.22	8,431	15,161	12,647	7.14	phosphatidylinositol 4-kinase alpha 1-like isoform X2
Core	scaffold377567				9,670	7.09	
Core	scaffold390489	maker-scaffold390489-exonerate_protein2genome-gene-0.40	6,856	10,960	10,253	6.80	translation initiation factor eIF-2B subunit delta-like isoform X1
Core	scaffold269378	maker-scaffold269378-exonerate_protein2genome-gene-0.20	2,737	3,373	2,929	6.60	putative ubiquitin-conjugating enzyme E2-binding protein
Core	scaffold300029	maker-scaffold300029-exonerate_protein2genome-gene-0.22	8,431	15,161	8,712	6.59	phosphatidylinositol 4-kinase alpha 1-like isoform X2
Core	scaffold379800				11,077	6.46	
Core	scaffold387913	maker-scaffold387913-exonerate_protein2genome-gene-0.18	10,567	12,606	10,923	6.25	ABC transporter G family member 1-like
Core	scaffold390292	maker-scaffold390292-exonerate_protein2genome-gene-0.58	15,210	19,017	18,392	6.20	autophagy-related protein 18g-like
Core	scaffold278747	maker-scaffold278747-exonerate_protein2genome-gene-0.42	6	1,526	407	6.18	Mammalian uncoordinated homology 13, domain 2
Core	scaffold390437	maker-scaffold390437-exonerate_protein2genome-gene-0.19	4,391	7,624	7,266	6.06	stromal 70 kDa heat shock-related protein, chloroplastic-like
Core	scaffold300029	maker-scaffold300029-exonerate_protein2genome-gene-0.31	5,753	7,865	7,601	5.97	phosphatidylinositol 4-kinase alpha 1-like isoform X1
Core	scaffold283046	maker-scaffold283046-exonerate_protein2genome-gene-0.0	10,839	12,999	11,759	5.87	SNF2 domain protein
Core	scaffold302647	maker-scaffold302647-exonerate_protein2genome-gene-0.20	5,767	16,099	8,743	5.69	acyl-CoA-binding domain-containing protein 4
Core	scaffold344675				4,854	5.59	
Core	scaffold390437	maker-scaffold390437-exonerate_protein2genome-gene-0.19	4,391	7,624	6,709	5.42	stromal 70 kDa heat shock-related protein, chloroplastic-like

Core	C21305877	maker-C21305877-exonerate_protein2genome-gene-0.0	1	5,578	5,327	5.06	calcium-dependent protein kinase 28-like
Core	C21306143	maker-C21306143-exonerate_protein2genome-gene-0.0	1,349	7,645	3,800	4.93	delta-1-pyrroline-5-carboxylate synthase
Core	C21306143	maker-C21306143-exonerate_protein2genome-gene-0.0	1,349	7,645	4,070	4.90	delta-1-pyrroline-5-carboxylate synthase
Front	scaffold388871				20,870	9.83	
Front	scaffold338938	maker-scaffold338938-exonerate_protein2genome-gene-0.23	1,561	14,812	7,647	9.73	G-type lectin S-receptor-like serine/threonine-protein kinase
Front	scaffold388908				9,746	9.33	
Front	scaffold388908	maker-scaffold388908-exonerate_protein2genome-gene-0.0	1,339	9,510	6,533	8.32	metallo-dependent phosphatase-like protein
Front	scaffold358045	maker-scaffold358045-exonerate_protein2genome-gene-0.0	1,945	9,475	2,856	8.05	protein ENHANCED DISEASE RESISTANCE 2-like
Front	scaffold269728	maker-scaffold269728-exonerate_protein2genome-gene-0.57	21,107	25,181	23,382	8.00	translation initiation factor 2, partial
Front	scaffold286111				422	7.79	
Front	scaffold283046				9,011	7.51	
Front	scaffold302647				827	7.39	
Front	scaffold302647				1,278	6.67	
Front	scaffold388871				13,415	6.66	
Front	scaffold390447	maker-scaffold390447-exonerate_protein2genome-gene-0.2	2,708	2,818	2,752	6.60	Zinc finger, CW-type
Front	scaffold379800				6,548	6.42	
Front	scaffold388871				8,474	6.36	

Front	scaffold379800				4,115	6.17	
Front	scaffold369012				1,444	5.44	
Front	scaffold390226	maker-scaffold390226-exonerate_protein2genome-gene-0.2	5,699	10,555	5,726	5.16	synaptotagmin-1-like
Shared							
Core	scaffold388871				21,721	14.03	
Front	scaffold388871				21,693	10.49	
Core	scaffold380929	maker-scaffold380929-exonerate_protein2genome-gene-0.84	8,523	12,489	9,635	6.65	proteasome subunit beta type-4-like
Front	scaffold380929	maker-scaffold380929-exonerate_protein2genome-gene-0.84	8,523	12,489	9,620	8.10	proteasome subunit beta type-4-like
Core	scaffold264195	maker-scaffold264195-exonerate_protein2genome-gene-0.20	6,856	9,960	8,862	17.76	probable alkaline/neutral invertase D
Front	scaffold264195	maker-scaffold264195-exonerate_protein2genome-gene-0.20	6,856	9,960	9,117	6.21	probable alkaline/neutral invertase D

Table S3.3. Selective sweeps detected in *core* and *front* populations with the G12 method. Non-annotated sweeps correspond entirely to upstream or downstream intergenic regions. Sweeps also found using SweeD are shown in grey. Genes with outlier SNPs are in bold (see Table 3.2 in the main text).

Group	Scaffold	Gene code	Gene start	Gene end	Sweep start	Sweep center	Sweep end	Sweep length	G12	G2/G1	Genes in sweep region (BLASTX annotation)
Exclusive											
Core	scaffold385996	maker-scaffold385996-exonerate_protein2genome-gene-0.14	3,287	9,793	7,994	8963	9257	1,264	0.48	0.02	beta-galactosidase 3-like
Core	scaffold389846				15,204	15,420	15,684	481	0.36	0.11	
Core	scaffold389846	maker-scaffold389846-exonerate_protein2genome-gene-0.0	12,840	14,865	13,114	13,340	13,775	662	0.36	0.07	ubiquitin-fold modifier-conjugating enzyme 1
Core	scaffold362323				11,572	11,777	12,101	530	0.32	0.21	
Core	scaffold254074				938	1,282	1,644	707	0.31	0.20	
Core	scaffold338938				17,696	18,768	19,506	1,811	0.29	0.06	
Core	scaffold387982				24,588	24,729	24,897	310	0.29	0.05	
Core	scaffold369102	maker-scaffold369102-exonerate_protein2genome-gene-0.0	1,855	11,818	7,340	7,566	7,781	442	0.28	0.14	phototropin-2
Core	scaffold338938	maker-scaffold338938-exonerate_protein2genome-gene-0.23	1,561	14,812	5,881	6,194	6,390	510	0.28	0.24	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290
Core	scaffold286111	maker-scaffold286111-exonerate_protein2genome-gene-0.0	6,342	7,736	7,349	7,622	7,892	544	0.26	0.15	anthocyanidin 3-O-glucosyltransferase 5-like
Core	scaffold341995				3,322	3,566	3,743	422	0.26	0.08	
Core	scaffold387982	maker-scaffold387982-exonerate_protein2genome-gene-0.29	17,307	23,329	21,446	21,718	22,099	654	0.26	0.09	ARF guanine-nucleotide exchange factor GNOM-like

Core	scaffold377567				18,724	18,919	19,669	946	0.24	0.09	
Core	scaffold362323				12,410	12,569	12,919	510	0.24	0.05	
Core	scaffold338938	maker-scaffold338938-exonerate_protein2genome-gene-0.19	45,493	47,542	45,229	45,666	45,882	6,54	0.24	0.28	RNA pseudouridine synthase 5
Core	scaffold390447				6,010	6,326	6,839	830	0.24	0.19	
Core	scaffold254074				6,549	6,682	6,878	330	0.24	0.11	
Core	scaffold389846				22,743	22,934	23,098	356	0.23	0.10	
Core	scaffold350078	maker-scaffold350078-exonerate_protein2genome-gene-0.18	14,318	16,542	15,172	15,676	16,385	1,214	0.23	0.17	receptor-like protein kinase THESEUS 1
Front	scaffold358045	maker-scaffold358045-exonerate_protein2genome-gene-0.0	1,945	9,475	2,531	3,033	3,637	1,107	0.55	0.19	protein ENHANCED DISEASE RESISTANCE 2-like
Front	**scaffold369012**	**maker-scaffold369012-exonerate_protein2genome-gene-0.0**	**4,653**	**7,878**	**5,087**	**5,336**	**6,151**	**1,065**	**0.48**	**0.50**	**high-affinity nitrate transporter 3.1-like**
Front	scaffold361799	maker-scaffold361799-exonerate_protein2genome-gene-0.0	1,234	9,988	4,451	5,309	5,843	1,393	0.48	0.27	alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase-like
Front	scaffold377567	maker-scaffold377567-exonerate_protein2genome-gene-0.0	10,499	17,745	14,795	15,071	16,347	1,553	0.45	0.03	probable ubiquitin-conjugating enzyme E2 23 isoform X1
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.70	17,125	19,859	18,115	19,232	19,821	1,707	0.44	0.13	putative vesicle-associated membrane protein 726
Front	scaffold388871	maker-scaffold388871-exonerate_protein2genome-gene-0.1	14,341	16,445	15,910	16,687	16,926	1,017	0.40	0.26	probable serine/threonine-protein kinase SIS8
Front	scaffold377567				8,700	9,471	10,375	1,676	0.39	0.14	
Front	scaffold283046				15,611	15,865	16,287	677	0.36	0.29	
Front	**scaffold283046**	**maker-scaffold283046-exonerate_protein2genome-gene-0.25**	**1,699**	**3,361**	**2,503**	**5,871**	**9,388**	**6,886**	**0.35**	**0.19**	**1-aminocyclopropane-1-carboxylate synthase 7-like**
Front	scaffold269719				30,026	31,329	32,792	2,767	0.35	0.13	

Front	scaffold390498				342	769	5,122	4,781	0.35	0.11	
Front	scaffold283046	maker-scaffold283046-exonerate_protein2genome-gene-0.0	10,839	12,999	11,034	12,810	13,831	2,798	0.33	0.43	SNF2 domain protein
Front	scaffold372370	maker-scaffold372370-exonerate_protein2genome-gene-0.0	4,594	6,676	4,401	4,726	5,157	757	0.33	0.48	vesicle transport protein, Use1
Front	**scaffold387203**	**maker-scaffold387203-exonerate_protein2genome-gene-0.0**	**290**	**1,857**	**511**	**1,047**	**2,124**	**1,614**	**0.32**	**0.57**	**high affinity nitrate transporter 2.7**
Front	scaffold379285	maker-scaffold379285-exonerate_protein2genome-gene-0.2	309	3,527	2,191	2,378	2,775	585	0.31	0.22	LRR receptor-like serine/threonine-protein kinase GSO1
Front	scaffold260786	maker-scaffold260786-exonerate_protein2genome-gene-0.1	4,007	7,956	3,683	3,976	4,441	759	0.31	0.10	cyclin-dependent kinase F-4-like isoform X2
Front	scaffold377567	maker-scaffold377567-exonerate_protein2genome-gene-0.0	10,499	17,745	17,029	18,273	18,489	1,461	0.31	0.17	probable ubiquitin-conjugating enzyme E2 23 isoform X1
Front	scaffold388871	maker-scaffold388871-exonerate_protein2genome-gene-0.1	14,341	16,445	14,221	14,782	15,049	829	0.31	0.24	probable serine/threonine-protein kinase SIS8
Front	scaffold390226				666	1,741	2,073	1,408	0.31	0.07	
Front	scaffold390364				4,866	5,303	6,172	1,307	0.30	0.20	
Front	scaffold379285	maker-scaffold379285-exonerate_protein2genome-gene-0.2	309	3,527	1,169	1,315	1,632	464	0.29	0.10	LRR receptor-like serine/threonine-protein kinase GSO1
Front	scaffold269728	maker-scaffold269728-exonerate_protein2genome-gene-0.57	21,107	25,181	23,052	23,751	24,393	1,342	0.29	0.42	translation initiation factor 2, partial
Front	scaffold384549				7,511	7,936	8,292	782	0.29	0.10	
Front	scaffold369102	maker-scaffold369102-exonerate_protein2genome-gene-0.0	1,855	11,818	5,557	5,834	6,160	604	0.29	0.14	phototropin-2
Front	scaffold284666				422	807	1,272	851	0.29	0.07	
Front	scaffold269378	maker-scaffold269378-exonerate_protein2genome-gene-0.0	11,166	12,440	11,718	12,003	12,372	655	0.28	0.32	peroxidase 72-like
Front	scaffold390226	maker-scaffold390226-exonerate_protein2genome-gene-0.0	4,273	4,713	3,163	4,014	4,318	1,156	0.28	0.19	putative ascorbate peroxidase, partial
Front	scaffold283046				591	1,154	1,667	1,077	0.28	0.26	

Front	scaffold384549				8,405	8,845	9,942	1,538	0.28	0.13	
Front	scaffold369102	maker-scaffold369102-exonerate_protein2genome-gene-0.0	1,855	11,818	8,975	9,463	9,839	865	0.28	0.15	phototropin-2
Front	scaffold361799	maker-scaffold361799-exonerate_protein2genome-gene-0.0	1,234	9,988	2,109	2,483	3,129	1,021	0.28	0.25	alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase-like
Front	scaffold236124				18,635	19,392	19,661	1,027	0.27	0.33	
Front	scaffold343087				1,409	1,645	2,012	604	0.26	0.20	
Front	scaffold369012				8,374	8,616	8,851	478	0.26	0.27	
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.11	6,632	9,762	7,621	8,243	8,628	1,008	0.26	0.32	ankyrin repeat domain-containing protein 13C-like
Front	scaffold381190	maker-scaffold381190-exonerate_protein2genome-gene-0.21	3,614	6,068	3,193	4,367	4,648	1,456	0.26	0.27	serine/threonine-protein kinase PBL27-like isoform X2
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.49	22,629	28,271	26,951	27,869	28,789	1,839	0.26	0.18	chromosome transmission fidelity protein 18 homolog
Front	C21305901	maker-C21305901-exonerate_protein2genome-gene-0.6	6,790	8,913	7,798	8,048	8,198	401	0.26	0.12	adenylyl-sulfate kinase 3-like isoform X1
Front	scaffold377567				23,443	23,659	24,054	612	0.26	0.10	
Front	scaffold283046	maker-scaffold283046-exonerate_protein2genome-gene-0.7	13,904	14,985	14,467	14,930	15,408	942	0.25	0.25	tetraspanin-2-like
Front	scaffold302647	maker-scaffold302647-exonerate_protein2genome-gene-0.0	1,553	4,003	2,189	2,646	3,110	922	0.25	0.55	pentatricopeptide repeat-containing protein At2g03880, mitochondrial-like
Front	scaffold254074				9,692	10,128	10,603	912	0.25	0.23	
Front	scaffold384549	maker-scaffold384549-exonerate_protein2genome-gene-0.0	10,072	12,573	11,605	11,797	12,033	429	0.25	0.37	protein WVD2-like 2
Front	scaffold385996	maker-scaffold385996-exonerate_protein2genome-gene-0.0	237	2,483	936	1,693	2,662	1,727	0.24	0.35	protein RTE1-HOMOLOG-like
Front	scaffold377678				4,041	4,265	4,401	361	0.24	0.39	
Front	scaffold377678				8,682	9,265	9,694	1,013	0.24	0.24	

Front	scaffold380929	maker-scaffold380929-exonerate_protein2genome-gene-0.47	17,460	22,510	20,282	20,721	21,218	937	0.24	0.56	probable protein phosphatase 2C 11
Front	scaffold390226	maker-scaffold390226-exonerate_protein2genome-gene-0.2	5,699	10,555	6,540	7,242	7,606	1,067	0.24	0.24	synaptotagmin-1-like
Front	scaffold296003	maker-scaffold296003-exonerate_protein2genome-gene-0.8	4,279	7,166	3,837	4,291	4,530	694	0.23	0.31	probable 2-carboxy-D-arabinitol-1-phosphatase
Front	scaffold388908	maker-scaffold388908-exonerate_protein2genome-gene-0.0	1,339	9,510	7,255	7,625	7,836	582	0.23	0.36	metallo-dependent phosphatase-like protein
Front	scaffold384549	maker-scaffold384549-exonerate_protein2genome-gene-0.0	10,072	12,573	10,814	11,057	11,475	662	0.23	0.21	protein WVD2-like 2
Front	scaffold379285	maker-scaffold379285-exonerate_protein2genome-gene-0.2	309	3,527	3,346	3,648	3,934	589	0.23	0.21	LRR receptor-like serine/threonine-protein kinase GSO1
Front	scaffold387913	maker-scaffold387913-exonerate_protein2genome-gene-0.1	1,909	4,582	3,701	4,028	4,626	926	0.23	0.29	serine/arginine-rich splicing factor RS2Z33-like isoform X1
Front	scaffold269719				20,575	20,952	21,495	921	0.23	0.16	
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.49	22,629	28,271	22,238	23,149	23,708	1,471	0.22	0.19	chromosome transmission fidelity protein 18 homolog
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.11	6,632	9,762	9,672	10,880	11,146	1,475	0.22	0.15	ankyrin repeat domain-containing protein 13C-like
Front	scaffold364572	maker-scaffold364572-exonerate_protein2genome-gene-0.0	5,532	7,895	7,546	7,988	8,576	1,031	0.22	0.12	histidine-containing phosphotransfer protein 1-like
Front	scaffold379800				4,618	5,061	5,434	817	0.22	0.25	
Front	scaffold390498				8,011	8,241	8,609	599	0.22	0.11	
Front	scaffold361799	maker-scaffold361799-exonerate_protein2genome-gene-0.0	1,234	9,988	6,858	7,088	7,673	816	0.22	0.55	alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase-like
Front	scaffold369102	maker-scaffold369102-exonerate_protein2genome-gene-0.0	1,855	11,818	8,004	8,404	8,856	853	0.22	0.16	phototropin-2
Front	scaffold338938	maker-scaffold338938-exonerate_protein2genome-gene-0.23	1,561	14,812	8,886	9,014	9,217	332	0.22	0.06	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.11	6,632	9,762	6,234	6,653	7,423	1,190	0.21	0.36	ankyrin repeat domain-containing protein 13C-like
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.30	12,089	15,839	12,592	13,188	13,548	957	0.21	0.22	protein SULFUR DEFICIENCY-INDUCED 2

Front	scaffold236124	maker-scaffold236124-exonerate_protein2genome-gene-0.0	10,546	14,006	10,574	11,473	11,683	1,110	0.21	0.40	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290
Front	scaffold380929				33,950	34,609	35,016	1,067	0.21	0.54	
Front	scaffold390437	maker-scaffold390437-exonerate_protein2genome-gene-0.19	4,391	7,624	5,911	6,219	6,537	627	0.21	0.39	stromal 70 kDa heat shock-related protein, chloroplastic-like
Front	scaffold302647				1,054	1,315	1,490	437	0.21	0.57	

Shared

Core	scaffold269378	maker-scaffold269378-exonerate_protein2genome-gene-0.17	3,458	5,199	5,552	5,751	6,357	806	0.43	0.11	putative ubiquitin-conjugating enzyme E2-binding protein
Front	scaffold269378	maker-scaffold269378-exonerate_protein2genome-gene-0.17	3,458	5,199	5,077	5,464	6,114	1,038	0.26	0.32	putative ubiquitin-conjugating enzyme E2-binding protein
Core	scaffold269378				7,048	7,486	7,752	705	0.27	0.08	
Front	scaffold269378				7,486	7,990	8,161	676	0.22	0.28	
Core	scaffold269378	maker-scaffold269378-exonerate_protein2genome-gene-0.23	13,124	16,468	14,961	15,923	16,754	1,794	0.38	0.07	probable ubiquitin-conjugating enzyme E2 16
Front	scaffold269378	maker-scaffold269378-exonerate_protein2genome-gene-0.23	13,124	16,468	12,470	13,786	16,590	4,121	0.45	0.07	probable ubiquitin-conjugating enzyme E2 16
Core	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.0	40	2,996	1,636	3,163	3,799	2,164	0.31	0.14	aspartate--tRNA ligase 2, cytoplasmic
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.0	40	2,996	2,184	2,565	3,027	844	0.26	0.23	aspartate--tRNA ligase 2, cytoplasmic
Core	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.87	34,181	36,849	33,816	34,462	35,722	1,907	0.68	0.01	putative F-box protein PP2-B12
Front	scaffold269719	maker-scaffold269719-exonerate_protein2genome-gene-0.87	34,181	36,849	33,933	34,777	36,704	2,772	0.50	0.04	putative F-box protein PP2-B12
Core	scaffold286111	maker-scaffold286111-exonerate_protein2genome-gene-0.40	8,492	11,253	10,105	10,484	10,781	677	0.23	0.25	arabinosyltransferase RRA3
Front	scaffold286111	maker-scaffold286111-exonerate_protein2genome-gene-0.40	8,492	11,253	9,344	10,099	10,456	1,113	0.21	0.47	arabinosyltransferase RRA3

Core	scaffold302647	maker-scaffold302647-exonerate_protein2genome-gene-0.20	5,767	16,099	6,800	7,178	7,853	1,054	0.39	0.11	acyl-CoA-binding domain-containing protein 4
Front	scaffold302647	maker-scaffold302647-exonerate_protein2genome-gene-0.20	5,767	16,099	6,473	6,855	7,181	709	0.29	0.24	acyl-CoA-binding domain-containing protein 4
Core	scaffold362323				628	993	2,093	1,466	0.34	0.13	
Front	scaffold362323				747	1,497	1,962	1,216	0.37	0.06	
Core	scaffold362323				5,110	5,394	7,226	2,117	0.33	0.10	
Front	scaffold362323				4,740	8,089	9,390	4,651	0.43	0.14	
Core	scaffold362323				10,243	10,367	10,581	339	0.34	0.06	
Front	scaffold362323				10,071	10,316	11,313	1,243	0.31	0.05	
Core	scaffold364572				2,117	2,295	2,506	390	0.26	0.08	
Front	scaffold364572				1,384	1,756	2,409	1,026	0.43	0.45	
Core	scaffold364572				3,330	3,439	3,583	254	0.23	0.15	
Front	scaffold364572				3,142	3,352	3,583	442	0.29	0.12	
Core	scaffold377567				4,635	4,881	5,093	459	0.23	0.18	
Front	scaffold377567				4,208	4,421	4,718	511	0.21	0.31	
Core	scaffold379800	maker-scaffold379800-exonerate_protein2genome-gene-0.0	7,006	9,653	6,775	7,281	8,105	1,331	0.39	0.30	receptor protein kinase CLAVATA1-like
Front	scaffold379800	maker-scaffold379800-exonerate_protein2genome-gene-0.0	7,006	9,653	6,625	6,945	7,583	959	0.42	0.32	receptor protein kinase CLAVATA1-like
Core	scaffold382282	maker-scaffold382282-exonerate_protein2genome-gene-0.17	9,506	15,366	9,679	9,943	10,206	528	0.31	0.06	dynamin-related protein 1E-like
Front	scaffold382282	maker-scaffold382282-exonerate_protein2genome-gene-0.17	9,506	15,366	9,786	9,969	10,275	490	0.24	0.15	dynamin-related protein 1E-like

Core	scaffold385996	maker-scaffold385996-exonerate_protein2genome-gene-0.14	3287	9793	4459	4962	5576	1118	0.36	0.30	beta-galactosidase 3-like
Front	scaffold385996	maker-scaffold385996-exonerate_protein2genome-gene-0.14	3287	9793	4378	4933	6005	1628	0.33	0.14	beta-galactosidase 3-like
Core	scaffold387913	maker-scaffold387913-exonerate_protein2genome-gene-0.18	10567	12606	11533	12357	12637	1105	0.30	0.13	ABC transporter G family member 1-like
Front	scaffold387913	maker-scaffold387913-exonerate_protein2genome-gene-0.18	10567	12606	10379	10770	12031	1653	0.36	0.04	ABC transporter G family member 1-like
Core	scaffold390226	maker-scaffold390226-exonerate_protein2genome-gene-0.2	5699	10555	8622	8833	9001	380	0.25	0.14	synaptotagmin-1-like
Front	scaffold390226	maker-scaffold390226-exonerate_protein2genome-gene-0.2	5699	10555	8982	9762	10125	1144	0.26	0.27	synaptotagmin-1-like

Table S3.4. Information for the 78 samples of *Leontodon longirostris* for which genotypic and phenotypic data was available for different half-sib plants. Flowering time refers to the number of days from seedling emergence to bolting (late winter common garden; de Pedro et al., unpublished). SNPs names are the same as in Table 3.2.

Population code	Maternal family	SNP S1 (position 2,041 bp)	SNP S2 (position 2,185 bp)	Genotype S1_S2	Flowering time (days)
05	p08	GG	GG	GGGG	33
05	p16	GG	GG	GGGG	49
05	p13	GG	GG	GGGG	53
05	p15	GG	GG	GGGG	56
05	p23	GG	GG	GGGG	73
10	p10	AG	GG	AGGG	49
10	p09	AG	GG	AGGG	52
10	p20	GG	GG	GGGG	47
10	p16	GG	GG	GGGG	52
23	p16	AG	AG	AGAG	47
23	p13	GG	GG	GGGG	42
23	p14	GG	GG	GGGG	42
23	p01	GG	GG	GGGG	48
25	p09	AG	AG	AGAG	26
25	p23	AG	AG	AGAG	47
25	p10	GG	GG	GGGG	50
27	p17	AG	AG	AGAG	56
27	p39	AG	AG	AGAG	57
35	p06	AA	AG	AAAG	52
35	p03	AG	AG	AGAG	45
35	p02	AG	AG	AGAG	47
35	p20	AG	GG	AGGG	50
35	p13	AG	GG	AGGG	80
35	p23	GG	AG	GGAG	61
35	p01	GG	GG	GGGG	43
35	p05	GG	GG	GGGG	50
35	p04	GG	GG	GGGG	52
35	p09	GG	GG	GGGG	63
55	p03	AG	AG	AGAG	54
55	p13	AG	AG	AGAG	70
55	p09	AG	GG	AGGG	56
55	p02	GG	GG	GGGG	41
55	p04	GG	GG	GGGG	53
55	p23	GG	GG	GGGG	54
55	p01	GG	GG	GGGG	68
55	p07	GG	GG	GGGG	78
60	p07	AG	GG	AGGG	47

60	p14	GG	GG	GGGG	44
60	p18	GG	GG	GGGG	50
63	p22	AA	AG	AAAG	50
63	p12	AG	AG	AGAG	56
63	p15	GG	GG	GGGG	45
63	p03	GG	GG	GGGG	50
63	p05	GG	GG	GGGG	71
67	p12	AA	AA	AAAA	57
67	p17	AA	AA	AAAA	66
67	p13	AA	AA	AAAA	68
67	p10	AA	AA	AAAA	71
67	p21	AA	AA	AAAA	79
67	p15	AA	AA	AAAA	89
67	p09	AG	AG	AGAG	48
67	p08	AG	AG	AGAG	58
67	p11	AG	AG	AGAG	60
68	p39	AG	AG	AGAG	53
68	p23	AG	AG	AGAG	60
68	p37	AG	AG	AGAG	73
68	p25	AG	AG	AGAG	79
69	p18	AG	AG	AGAG	49
69	p09	AG	AG	AGAG	51
72	p22	AA	AA	AAAA	73
72	p13	AA	AA	AAAA	93
72	p21	AG	AG	AGAG	77
74	p19	AA	AA	AAAA	121
74	p12	AG	AG	AGAG	53
74	p15	AG	AG	AGAG	79
74	p14	AG	AG	AGAG	115
74	p08	GG	GG	GGGG	66
74	p25	GG	GG	GGGG	86
78	p08	AA	AA	AAAA	66
78	p13	AA	AA	AAAA	75
78	p02	AA	AA	AAAA	76
78	p17	AA	AA	AAAA	86
78	p20	AA	AA	AAAA	89
78	p25	AA	AA	AAAA	111
78	p24	AG	AG	AGAG	60
78	p07	AG	AG	AGAG	77
78	p06	AG	AG	AGAG	113
78	p01	GG	GG	GGGG	97

References

A

Agrawal, A. A., Hastings, A. P., Bradburd, G. S., Woods, E. C., Züst, T., Harvey, J. A., & Bukovinszky, T. (2015). Evolution of plant growth and defense in a continental introduction. *American Naturalist*, 186, E1-E15.

Alexander, J. M., & Edwards, P. J. (2010). Limits to the niche and range margins of alien species. *Oikos*, 119, 1377-1386.

Alleaume-Benharira, M., Pen, I. R., & Ronce, O. (2006). Geographical patterns of adaptation within a species' range: Interactions between drift and gene flow. *Journal of Evolutionary Biology*, 19, 203-215.

Álvarez-Lao, D. J., & Méndez, M. (2016). Latitudinal gradients and indicator species in ungulate paleoassemblages during the MIS 3 in W Europe. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 449, 455-462.

Amundsen, P. A., Salonen, E., Niva, T., Gjelland, K. Ø., Præbel, K., Sandlund, O. T., … Bøhn, T. (2012). Invader population speeds up life history during colonization. *Biological Invasions*, 14, 1501-1513.

Arc, E., Sechet, J., Corbineau, F., Rajjou, L., & Marion-Poll, A. (2013). ABA crosstalk with ethylene and nitric oxide in seed dormancy and germination. *Frontiers in Plant Science*, 4, 63.

Aronson, J., Kigel, J., & Shmida, A. (1993). Reproductive allocation strategies in desert and mediterranean populations of annual plants grown with and without water stress. *Oecologia*, 93, 336-342.

Aronson, J., Kigel, J., Shmida, A., & Klein, J. (1992). Adaptive phenology of desert and Mediterranean populations of annual plants grown with and without water stress. *Oecologia*, 89, 17-26.

Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P. H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, 51, 148-164.

B

Baker, H. G. (1955). Self-compatibility and establishment after "long-distance" dispersal. *Evolution*, 9, 347-349.

Barrett, S. C. H. (2014). Evolution of mating systems: outcrossing versus selfing. In J. Losos (Ed.), *The Princeton Guide to Evolution.* (pp. 356-362). Princeton: Princeton University Press.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23, 38-44.

Bates, D., Mächler, M., Bolker, B, & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67, 1-48.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., … Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, 157, 80-97.

Bitter, M. C., Kapsenberg, L., Gattuso, J.-P., & Pfister, C. A. (2019). Standing genetic variation fuels rapid adaptation to ocean acidification. *Nature Communications*, 10, 5821.

Blows, M. W., & Hoffmann, A. A. (2005). A reassessment of genetic limits to evolutionary change. *Ecology*, 86, 1371-1384.

Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., Corbeil, J. (2012). Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biology*, 13, R122.

Bossdorf, O., Auge, H., Lafuma, L., Rogers, W. E., Siemann, E., & Prati, D. (2005). Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, 144, 1-11.

Bosshard, L., Dupanloup, I., Tenaillon, O., Bruggmann, R., Ackermann, M., Peischl, S., & Excoffier, L. (2017). Accumulation of deleterious mutations during bacterial range expansions. *Genetics*, 207, 669-684.

Bosshard, L., Peischl, S., Ackermann, M. & Excoffier, L. (2019). Mutational and selective processes involved in evolution during bacterial range expansions. *Molecular Biology and Evolution*, 36, 2313-2327.

Brachi, B., Aimé, C., Glorieux, C., Cuguen, J., & Roux, F. (2012). Adaptive value of phenological traits in stressful environments: Predictions based on seed production and laboratory natural selection. *PLoS ONE*, 7, e32069.

Braendle, C., Heyland, A. & Flatt, T. (2011). Integrating mechanistic and evolutionary analysis of life history variation. In T. Flatt and A. Heyland (Eds.), *Mechanisms of life history evolution: the genetics and physiology of life history traits and trade-off*. (pp. 3-10). Oxford: Oxford University Press.

Brändel, M. (2007). Ecology of achene dimorphism in *Leontodon saxatilis*. *Annals of Botany*, 100, 1189-1197.

Brennan, R. S., Garrett, A. D., Huber, K. E., Hargarten, H., & Pespeni, M. H. (2019). Rare genetic variation and balanced polymorphisms are important for survival in global change conditions. *Proceedings of the Royal Society B*, 286, 20190943.

Bridle, J. R., & Vines, T. H. (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, 22, 140-147.

Brockmann, D., & Helbing, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342, 1337-1342.

Brouillette, L. C., Mason, C. M., Shirk, R. Y., & Donovan, L. A. (2014). Adaptive differentiation of traits related to resource use in a desert annual along a resource gradient. *New Phytologist*, 201, 1316-1327.

Burghardt, L. T., Metcalf, C. J. E., Wilczek, A. M., Schmitt, J., & Donohue, K. (2015). Modeling the influence of genetic and environmental variation on the expression of plant life cycles across landscapes. *American Naturalist*, 185, 212-227.

Bürstenbinder, K., Rzewuski, G., Wirtz, M., Hell, R., & Sauter, M. (2007). The role of methionine recycling for ethylene synthesis in Arabidopsis. *The Plant Journal*, 49, 238-249.

Burton, O. J., Phillips, B. L., & Travis, J. M. J. (2010). Trade-offs and the evolution of life-histories during range expansion. *Ecology Letters*, 13, 1210-1220.

C

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., & Yandell. M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18, 188-196.

Castro, J. (2006). Short delay in timing of emergence determines establishment success in *Pinus sylvestris* across microhabitats. *Annals of Botany*, 98, 1233-1240.

CGP: The Composite Genome Project (2020, July 8). *Priorities for research, education and extension in genomics, genetics, and breeding of the Compositae* [White paper]. Retrieved from The Composite Genome Project, University of California, Davis: https://compgenomics.ucdavis.edu/cwp/white_paper_draft_09_12_07.pdf

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10, 195-205.

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134, 1289-1303.

Chen, I., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species of climate warming. *Science*, 333, 1024-1026.

Chiang, G. C. K., Barua, D., Kramer, E. M., Amasino, R. M., & Donohue, K. (2009). Major flowering time gene, *FLOWERING LOCUS FLC*, regulates seed germination in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 11661-1666.

Chopin, F., Orsel, M., Dorbe, M.-F., Chardon, F., Truong, H.-N., Miller, A. J., … Daniel-Vedelea, F. (2007). The Arabidopsis ATNRT2.7 nitrate transporter controls nitrate content in seeds. *The Plant Cell*, 19, 1590-1602.

Chuang, A., & Peterson, C. R. (2016). Expanding population edges: theories, traits, and trade-offs. *Global Change Biology*, 22, 494-512.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; iso-2; iso-3. *Fly*, 6, 80-92.

Claros, M. G., Seoane, P., Bautista, R., González-Martínez, S., Riba, M., Mayol, M., & de Pedro, M. (2020). *Leontodon longirostris genome draft 418Mb*. figshare. Dataset. https://doi.org/10.6084/m9.figshare.12706247.v3

Cohen, D. (1966). Optimizing reproduction in a randomly varying environment. *Journal of Theoretical Biology*, 12, 119-129.

Colautti, R. I., Alexander, J. M., Dlugosch, K. M., Keller, S. R., & Sultan, S. E. (2017). Invasions and extinctions through the looking glass of evolutionary ecology. *Philosophical Transactions of the Royal Society B*, 372, 20160031.

Colautti, R. I., & Barrett, S. C. H. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, 342, 364-366.

Colautti, R. I., Eckert, C. G., & Barrett, S. C. H. (2010). Evolutionary constraints on adaptive evolution during range expansion in an invasive plant. *Proceedings of the Royal Society B: Biological Sciences*, 277, 1799-1806.

Colautti, R. I., & Lau, J. A. (2015). Contemporary evolution during invasion: Evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, 24, 1999-2017.

Cotto, O., & Ronce, O. (2014). Maladaptation as a source of senescence in habitats variable in space and time. *Evolution*, 68, 2481-2493.

Crisci, J. L., Poh, Y. P., Mahajan, S., & Jensen, J. D. (2013). The impact of equilibrium assumptions on tests of selection. *Frontiers in Genetics*, 4, 235.

Crow, J. F., & Kimura, M. (1970). *An introduction to population genetic theory*. New York: Harper and Row.

Cwynar, L. C., & MacDonald, G. M. (1987). Geographical variation of lodgepole pine in relation to population history. *American Naturalist*, 129, 463-469.

D

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156-2158.

Davis, M. B., & Shaw, R. G. (2001). range shifts and adaptive responses to quaternary climate change. *Science*, 292, 673-679.

Davis, M. B., Shaw, R. G., & Etterson, J. R. (2005). Evolutionary responses to changing climate. *Ecology*, 86, 1704-1714.

de Castro, M., Martín-Vide, J., & Alonso, S. (2005). *The climate of Spain: past, present and scenarios for the 21st century. A preliminary assessment of the impacts in Spain due to the effects of climate change*. ECCE Project-Final report. Madrid: Ministerio de Medio Ambiente.

De Meester, L., Stoks, R., & Brans, K. I. (2018). Genetic adaptation as a biological buffer against climate change: Potential and limitations. *Integrative Zoology*, 13, 372-391.

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, 22, 1383-1399.

de Pedro, M., Riba, M., González-Martínez, S. C., Seoane, P., Bautista, R., Claros, M. G., & Mayol, M. (2021). Demography, genetic diversity and expansion load in the colonizing species *Leontodon longirostris* (Asteraceae) throughout its native range. *Molecular Ecology*, DOI: 10.1111/mec.15802.

Debieu, M., Tang, C., Stich, B., Sikosek, T., Effgen, S., Josephs, E., … de Meaux, J. (2013). Co-variation between seed dormancy, growth rate and flowering time changes with latitude in *Arabidopsis thaliana*. *PLoS ONE*, 8, e61075.

Del Fabbro, C., Scalabrin, S., Morgante, M., & Giorgi, F. M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS ONE*, 8, e85024.

Denisov, S. V., Bazykin, G. A., Sutormin, R., Favorov, A. V., Mironov, A. A., Gelfand, M. S., & Kondrashov, A. S. (2014). Weak negative and positive selection and the drift load at splice sites. *Genome Biology and Evolution*, 6, 1437-1447.

Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, 24, 2095-2111.

Donohue, K. (2002). Germination timing influences natural selection on life-history characters in *Arabidopsis thaliana*. *Ecology*, 83, 1006-1016.

Donohue, K. (2005). Niche construction through phenological plasticity: life history dynamics and ecological consequences. *New Phytologist*, 166, 83-92.

Donohue, K. (2009). Completing the cycle: Maternal effects as the missing link in plant life histories. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1059-1074.

Donohue, K., Rubio De Casas, R., Burghardt, L., Kovach, K., & Willis, C. G. (2010). Germination, postgermination adaptation, and species ecological ranges. *Annual Review of Ecology, Evolution, and Systematics*, 41, 293-319.

Dornier, A., Munoz, F., & Cheptou, P.-O. (2008). Allee effect and self-fertilization in hermaphrodites: reproductive assurance in a structured metapopulation. *Evolution*, 62, 2558-2569.

Dyer, A. R., Fenech, A., & Rice, K. J. (2000). Accelerated seedling emergence in interspecific competitive neighbourhoods. *Ecology Letters*, 3, 523-529.

E

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Molecular Ecology*, 17, 1170-1188.

Edmonds, C. A., Lillie, A. S., & Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 975-979.

Elam, D. R., Ridley, C. E., Goodell, K., & Ellstrand, N. C. (2007). Population size and relatedness affect fitness of a self-incompatible invasive plant. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 549-552.

Endler, J. A. (1977). *Geographic variation, speciation, and clines*. Princeton: Princeton University Press.

Epanchin-Niell, R. S., & Hastings, A. (2010). Controlling established invaders: Integrating economics and spread dynamics to determine optimal management. *Ecology Letters*, 13, 528-541.

Estoup, A., Beaumont, M., Sennedot, F., Moritz, C., & Cornuet, J. M. (2004). Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, 58, 2021-2036.

Etterson, J. R., & Shaw, R. G. (2001). Constraint to adaptive evolution in response to global warming. *Science*, 294, 151-154.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9, e1003905.

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481-501.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 10, 564-567.

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, 23, 347-351.

F

Falgueras, J., Lara, A. J., Fernández-Pozo, N., Cantón, F. R., Pérez-Trabado, G. & Claros, M. G. (2010). SeqTrim: a high-throughput pipeline forpre-processing any type of sequence read. *BMC Bioinformatics*, 11, 38.

Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2016). Going global by adapting local: A review of recent human adaptation. *Science*, 354, 54-59.

Franks, S. J., & Weis, A. E. (2008). A change in climate causes rapid evolution of multiple life-history traits and their interactions in an annual plant. *Journal of Evolutionary Biology*, 21, 1321-1334.

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, 46, 493-505.

Friedman, J. (2020). The evolution of annual and perennial plant life histories: Ecological correlates and genetic mechanisms. *Annual Review of Ecology, Evolution, and Systematics*, 51, 461-481.

Fronhofer, E. A., & Altermatt, F. (2015). Eco-evolutionary feedbacks during experimental range expansions. *Nature Communications*, 6:6844.

Fu, W., Gittelman, R. M., Bamshad, M. J., & Akey, J.M. (2014). Characteristics of neutral and deleterious protein-coding variation among individuals and populations. *The American Journal of Human Genetics*, 95, 421-436.

G

Galloway, L. F., & Burgess, K. S. (2009). Manipulation of flowering time: Phenological integration and maternal effects. *Ecology*, 90, 2139-2148.

Galloway, L. F., Watson, R. H. B., & Prendeville, H. R. (2018). Response to joint selection on germination and flowering phenology depends on the direction of selection. *Ecology and Evolution*, 8, 7688-7696.

García, H. (2004). *Consecuencias ecológicas de la sucesión sobre algunas características poblacionales e individuales de Leontodon taraxacoides (Vill.) Mérat* (Unpublished Master's Thesis). Autonomous University of Barcelona, Barcelona, Spain.

Garud, N. R., & Petrov, D. A. (2016). Elevated linkage disequilibrium and signatures of soft sweeps are common in *Drosophila melanogaster*. *Genetics*, 203, 863-880.

Gaston, K. J. (2003). *The structure and dynamics of geographic ranges*. Oxford: Oxford University Press.

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, 201, 1555-1579.

Gilbert, K. J., Sharp, N. P., Angert, A. L., Conte, G. L., Draghi, J. A., Guillaume, F., … Whitlock, M. C. (2017). Local adaptation interacts with expansion load during range expansion: maladaptation reduces expansion load. *American Naturalist*, 189, 368-380.

Gómez, A., & Lunt, D. H. (2007). Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In S. Weiss and N. Ferrand (Eds.), *Phylogeography of Southern European Refugia*. (pp. 155-188). Dordrecht: Springer.

González-Martínez, S. C., Ridout, K., & Pannell, J. R. (2017). Range expansion compromises adaptive evolution in an outcrossing plant. *Current Biology*, 27, 2544-2551.

González-Sampériz, P., Leroy, S. A. G., Carrión, J. S., Fernández, S., García-Antón, M., Gil-García, M. J., … Figueiral, I. (2010). Steppes, savannahs, forests and phytodiversity reservoirs during the Pleistocene in the Iberian Peninsula. *Review of Palaeobotany and Palynology*, 162, 427-457.

Gralka, M., Stiewe, F., Farrell, F., Möbius, W., Waclaw, B., & Hallatschek, O. (2016). Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters*, 19, 889-898.

Gravel, S. (2016). When is selection effective? *Genetics*, 203, 451-462.

Griffith, T. M., & Watson, M. A. (2005). Stress avoidance in a common annual: Reproductive timing is important for local adaptation and geographic distribution. *Journal of Evolutionary Biology*, 18, 1601-1612.

Griffith, T. M., & Watson, M. A. (2006). Is evolution necessary for range expansion? Manipulating reproductive timing of a weedy annual transplanted beyond its range. *American Naturalist*, 167, 153-164.

Groom, M. J. (1998). Allee effects limit population viability of an annual plant. *American Naturalist*, 151, 487-496.

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications*, 11, 1001.

Groves, R. H., Hosking, J. R., Batianoff, G. N., Cooke, D. A., Cowie, I. D., Johnson, R. W., … Waterhouse, B. M. (2003). *Weed Categories for Natural and Agricultural Ecosystem Management*. Australian Government: Department of Agriculture, Fisheries and Forestry.

H

Hallatschek, O., & Nelson, D. R. (2008). Gene surfing in expanding populations. *Theoretical Population Biology*, 73, 158-170.

Halliday, K. J., & Whitelam, G. C. (2003). Changes in photoperiod or temperature alter the functional relationships between phytochromes and reveal roles for *phyD* and *phyE*. *Plant Physiology*, 131, 1913-1920.

Hamilton, W. D., & May, R. M. (1977). Dispersal in stable habitats. *Nature*, 269, 578-581.

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., … Bergelson, J. (2011). *Arabidopsis thaliana* Genome. *Science*, 334, 83-86.

Harpak, A., Garud, N., Rosenberg, N. A., Petrov, D. A., Combs, M., Pennings, P. S., & Munshi-South, J. (2021). Genetic adaptation in New York city rats. *Genome Biology and Evolution*, 13, evaa247.

Harris, A. M., Garud, N. R., & DeGiorgio, M. (2018). Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. *Genetics*, 210, 1429-1452.

Harris, R. B., Sackman, A., & Jensen, J. D. (2018). On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. *PLoS Genetics*, 14, e1007859.

Hartfield, M., & Otto, S. P. (2011). Recombination and hitchhiking of deleterious alleles. *Evolution*, 65, 2421-2434.

Henn, B. M., Botigué, L. R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B. K., … Bustamante, C. D. (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 113, E440-E449.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *American Naturalist*, 173, 579-588.

Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169, 2335-2352.

Heschel, M. S., Selby, J., Butler, C., Whitelam, G. C., Sharrock, R. A., & Donohue, K. (2007). A new role for phytochromes in temperature-dependent germination. *New Phytologist*, 174, 735-741.

Hewitt, G. M. (1999). Postglacial re-colonisation of European biota. *Biological Journal of the Linnean Society*, 68, 87-112.

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B*, 359, 183-195.

Hodgins, K. A., Lai, Z., Oliveira, L. O., Still, D. W., Scascitelli, M., Barker, M. S., … Rieseberg, L. H. (2014). Genomics of compositae crops: Reference transcriptome assemblies and evidence of hybridization with wild relatives. *Molecular Ecology Resources*, 14, 166-177.

Hoffmann, A. a, & Sgrò, C. M. (2011). Climate change and evolutionary adaptation. *Nature*, 470, 479-485.

Holt, R. D., & Barfield, M. (2011). Theoretical perspectives on the statics and dynamics of species' borders in patchy environments. *American Naturalist*, 178, S6-S25.

Huang, F., Peng, S., Chen, B., Liao, H., Huang, Q., Lin, Z., & Liu, G. (2015). Rapid evolution of dispersal-related traits during range expansion of an invasive vine *Mikania micrantha*. *Oikos*, 124, 1023-1030.

Hunt, R. C., Simhadri, V. L., Iandoli, M., Sauna, Z. E., & Kimchi-Sarfaty, C. (2014). Exposing synonymous mutations. *Trends in Genetics*, 30, 308-321.

Huo, H., Wei, S., & Bradford, K. J. (2016). *DELAY OF GERMINATION1* (*DOG1*) regulates both seed dormancy and flowering time through microRNA pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 113, E2199–E2206.

I

Imaizumi, T., Auge, G., & Donohue, K. (2017). Photoperiod throughout the maternal life cycle, not photoperiod during seed imbibition, influences germination in *Arabidopsis thaliana*. *American Journal of Botany*, 104, 516-526.

J

Jiménez-Moreno, G., Fauquette, S., & Suc, J. P. (2010). Miocene to Pliocene vegetation reconstruction and climate estimates in the Iberian Peninsula from pollen data. *Review of Palaeobotany and Palynology*, 162, 403-415.

Joshi, J., & Vrieling, K. (2005). The enemy release and EICA hypothesis revisited: Incorporating the fundamental difference between specialist and generalist herbivores. *Ecology Letters*, 8, 704-714.

Just, A., Gourvil, J., Millet, J., Boullet, V., Milon, T., Mandon, I., & Dutrève, B. (2015). SIFlore, a dataset of geographical distribution of vascular plants covering five centuries of knowledge in France: Results of a collaborative project coordinated by the Federation of the National Botanical Conservatories. *PhytoKeys*, 56, 47-60.

K

Kalisz, S., & Kramer, E. M. (2008). Variation and constraint in plant evolution and development. *Heredity*, 100, 171-177.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, 7, 1225-1241.

Keane, R. M., & Crawley, M. J. (2002). Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution*, 17, 164-170.

Keller, S. R., Sowell, D. R., Neiman, M., Wolfe, L. M., & Taylor, D. R. (2009). Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytologist*, 183, 678-690.

Kiefer, C., Severing, E., Karl, R., Bergonzi, S., Koch, M., Tresch, A., & Coupland, G. (2017). Divergence of annual and perennial species in the Brassicaceae and the contribution of cis-acting variation at *FLC* orthologues. *Molecular Ecology*, 26, 3437-3457.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217, 624-626.

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a species' range. *American Naturalist*, 150, 1-23.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23, 482-490.

Kooyers, N. J., Greenlee, A. B., Colicchio, J. M., Oh, M., & Blackman, B. K. (2015). Replicate altitudinal clines reveal that evolutionary flexibility underlies adaptation to drought stress in annual *Mimulus guttatus*. *New Phytologist*, 206, 152-165.

Koski, M. H., Layman, N. C., Prior, C. J., Busch, J. W., & Galloway, L. F. (2019). Selfing ability and drift load evolve with range expansion. *Evolution Letters*, 3, 500-512.

Kronholm, I., Picó, F. X., Alonso-Blanco, C., Goudet, J., & de Meaux J. (2012). Genetic basis of adaptation in *Arabidopsis thaliana*: local adaptation at the seed dormancy QTL *DOG1*. *Evolution*, 66, 2287-2302.

Kryvokhyzha, D., Salcedo, A., Eriksson, M. C., Duan, T., Tawari, N., Chen, J., … Lascoux, M. (2019). Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid *Capsella bursa-pastoris* (Brassicaceae). *PLoS Genetics*, 15, e1007949.

Kurze, S., Bareither, N., & Metz, J. (2017). Phenology, roots and reproductive allocation, but not the LHS scheme, shape ecotypes along an aridity gradient. *Perspectives in Plant Ecology, Evolution and Systematics*, 29, 20-29.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82, DOI: 10.18637/jss.v082.i13.

L

Laenen, B., Tedder, A., Nowak, M. D., Toräng, P., Wunder, J., Wötzel, S., … Slotte, T. (2018). Demography and mating system shape the genome-wide impact of purifying selection in Arabis alpina. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 816-821.

Lai, Y.-T., Yeung, C. K. L., Omland, K. E., Pang, E.-L., Hao, Y., Liao, B.-Y., … Li, S.-H. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 2152-2157.

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, 37, 1210-1226.

Lande, R., & Schemske, D. W. (1985). The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution*, 39, 24-40.

Laughlin, D. C., & Messier, J. (2015). Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends in Ecology & Evolution*, 30, 487-496.

Lavergne, S., Hampe, A., Arroyo, J. (2013). In and out of Africa: how did the Strait of Gibraltar affect plant species migration and local diversification? *Journal of Biogeography*, 40, 24-36.

Lee, S. D., Park, S., Park, Y. S., Chung, Y. J., Lee, B. Y., & Chon, T. S. (2007). Range expansion of forest pest populations by using the lattice model. *Ecological Modelling*, 203, 157-166.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, 3, e4010.

Leishman, M. R., Haslehurst, T., Ares, A., & Baruch, Z. (2007). Leaf trait relationships of native and invasive plants: Community- and global-scale comparisons. *New Phytologist*, 176, 635-643.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17, 183-189.

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.2. https://CRAN.R-project.org/package=emmeans

Lewontin, R. C. (1965). Comment. In H. G. Baker and G. L. Stebbins (Eds.), *The genetics of colonizing species*. (pp. 481). New York: Academic Press.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., … Wang, J. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, 20, 265-272.

Linhart, Y. B., & Grant, M. C. (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology, and Systematics*, 27, 237-277.

Liu, A., & Burke, J. M. (2006). Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics*, 173, 321-330.

Liu, X., & Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*, 47, 555-559.

Liu, B., Yuan, J., Yiu, S. M., Li, Z., Xie, Y., Chen, Y., … Luo, R. (2012). COPE: An accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. *Bioinformatics*, 28, 2870-2874.

Lohmueller, K. E. (2014). The distribution of deleterious genetic variation in human populations. *Current Opinion in Genetics & Development*, 29, 139-146.

Lohmueller, K. E., Indap, A. R., Schmidt, S., Boyko, A. R, Hernandez, R. D., Hubisz, M. J., … Bustamante, C. D. (2008). Proportionally more deleterious genetic variation in European than in African populations. *Nature*, 451, 994-997.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1, 18.

Lustenhouwer, N., Wilschut, R. A., Williams, J. L., van der Putten, W. H., & Levine, J. M. (2018). Rapid evolution of phenology during range expansion with recent climate change. *Global Change Biology*, 24, e534-e544.

M

Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M., & Bazzaz, F. A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, 10, 689-710.

Manzano-Piedras, E., Marcer, A., Alonso-Blanco, C., & Picó, F. X. (2014). Deciphering the adjustment between environment and life history in annuals: Lessons from a geographically-explicit approach in *Arabidopsis thaliana*. *PLoS ONE*, 9, e87836.

Marcer, A., Vidigal, D. S., James, P. M. A., Fortin, M.-J., Méndez-Vigo, B., Hilhorst, H. W. M., … Picó, F. X. (2017). Temperature fine-tunes Mediterranean *Arabidopsis thaliana* life-cycle phenology geographically. *Plant Biology*, 38, 42-49.

Marks, M., & Prince, S. (1981). Influence of germination date on survival and fecundity in wild lettuce *Lactuca Serriola*. *Oikos*, 36, 326-330.

Marsden, C. D., Vecchyo, D. O. Del, O'Brien, D. P., Taylor, J. F., Ramirez, O., Vilà, C., … Lohmueller, K. E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 152-157.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, 10-12.

Martínez-Berdeja, A., Stitzer, M. C., Taylor, M. A., Okada, M., Ezcurra, E., Runcie, D. E., & Schmitt, J. (2020). Functional variants of *DOG1* control seed chilling responses and variation in seasonal life-history strategies in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 2526-2534.

Matakiadis, Th., Alboresi, A., Jikumaru, Y., Tatematsu, K., Pichon, O., Renou, J.-P., … Truong, H.-N. (2009). The Arabidopsis abscisic acid catabolic gene CYP707A2 plays a key role in nitrate control of seed dormancy. *Plant Physiology*, 149, 949-960.

Mattila, T. M., Aalto, E. A., Toivainen, T., Niittyvuopio, A., Piltonen, S., Kuittinen, H., & Savolainen, O. (2016). Selection for population-specific adaptation shaped

patterns of variation in the photoperiod pathway genes in *Arabidopsis lyrata* during post-glacial colonization. *Molecular Ecology*, 25, 581-597.

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, 23, 23-35.

Mayr, E. (1956). Geographical character gradients and climatic adaptation. *Evolution*, 10, 105-108.

McCarty J. P., Wolfenbarger, L. L. & Wilson, J. A. (2017). *Biological Impacts of Climate Change*. Encyclopedia of Life Sciences (ELS). Chichester: John Wiley & Sons.

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21, 2839-2846.

Méndez-Vigo, B., Gomaa, N. H., Alonso-Blanco, C., & Picó, F. X. (2013). Among- and within-population variation in flowering time of Iberian *Arabidopsis thaliana* estimated in field and glasshouse conditions. *New Phytologist*, 197, 1332-1343.

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution*, 28, 659-669.

Metcalf, J. C., Rose, K. E., & Rees, M. (2003). Evolutionary demography of monocarpic perennials. *Trends in Ecology & Evolution*, 18, 471-480.

Meusel, H., & Jäger, E. (1992). *Vergleichende Chorologie der zentraleuropäischen Flora, vol. III*. Jena: Gustav Fischer Verlag. http://chorologie.biologie.uni-halle.de/choro/index.php?Lang=E

Miller, T. E. X., Angert, A. L., Brown, C. D., Lee-Yaw, J. A., Lewis, M., Lutscher, F., … Williams, J. L. (2020). Eco-evolutionary dynamics of range expansion. *Ecology*, 101, e03139.

Montague, J. L., Barrett, S. C. H., & Eckert, C. G. (2008). Re-establishment of clinal variation in flowering time among introduced populations of purple loosestrife (*Lythrum salicaria*, Lythraceae). *Journal of Evolutionary Biology*, 21, 234-245.

Montesinos-Navarro, A., Picó, F. X., & Tonsor, S. J. (2012). Clinal variation in seed traits influencing life cycle timing in *Arabidopsis thaliana*. *Evolution*, 66, 3417-3431.

Mousseau, T. A., & Fox, C. W. (1998). The adaptive significance of maternal effects. *Trends in Ecology & Evolution*, 13, 403-407.

Müller-Schärer, H., Schaffner, U., & Steinger, T. (2004). Evolution in invasive plants: Implications for biological control. *Trends in Ecology & Evolution*, 19, 417-422.

Murren, C. J. (2012). The integrated phenotype. *Integrative and Comparative Biology*, 52, 64-76.

Mustin, K., Benton, T. G., Dytham, C., & Travis, J. M. J. (2009). The dynamics of climate-induced range shifting; perspectives from simulation modelling. *Oikos*, 118, 131-137.

N

Nadeau, C. P., & Urban, M. C. (2019). Eco-evolution on the edge during climate change. *Ecography*, 42, 1280-1297.

Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, 13, 1-6.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197-218.

Ninyerola, M., Pons, X., & Roure, J. M. (2005). *Atlas Climático Digital de la Península Ibérica. Metodología y aplicaciones en bioclimatología y geobotánica*. Bellaterra: Universidad Autónoma De Barcelona. ISBN 932860-8-7.

O

Ochocki, B. M., & Miller, T. E. X. (2017). Rapid evolution of dispersal ability makes biological invasions faster and more variable. *Nature Communications*, 8, 14315.

Ochocki, B. M., Saltz, J. B., & Miller, T. E. X. (2020). Demography-dispersal trait correlations modify the eco-evolutionary dynamics of range expansion. *American Naturalist*, 195, 231-246.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology, Evolution, and Systematics*, 23, 263-286.

Olsson, K., & Ågren, J. (2002). Latitudinal population differentiation in phenology, life history and flower morphology in the perennial herb *Lythrum salicaria*. *Journal of Evolutionary Biology*, 15, 983-996.

Ortiz, M. A., Tremetsberger, K., Stuessy, T. F., Terrab, A., García-Castaño, J. L., & Talavera, S. (2009). Phylogeographic patterns in *Hypochaeris* section *Hypochaeris* (Asteraceae, Lactuceae) of the western Mediterranean. *Journal of Biogeography*, 36, 1384-1397.

Otto, S. P., & Whitlock, M. C. (1997). The probability of fixation in populations of changing size. *Genetics*, 146, 723-733.

P

Pannell, J. R., & Barrett, S. C. H. (1998). Baker's Law revisited: reproductive assurance in a metapopulation. *Evolution*, 52, 657-668.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution, and Systematics*, 37, 637-669.

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37-42.

Pattyn, J., Vaughan-Hirsch, J., & Van de Poel, B. (2021). The regulation of ethylene biosynthesis: a complex multilevel control circuitry. *New Phytologist*, 229, 770-782.

Pavlidis, P., Jensen, J. D., & Stephan, W. (2010). Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*, 185, 907-922.

Pavlidis, P., Zivkovic, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, 30, 2224-2234.

Peiman, K. S., & Robinson, B. W. (2017). Comparative analyses of phenotypic trait covariation within and among populations. *American Naturalist*, 190, 451-468.

Peischl, S., Dupanloup, I., Bosshard, L., & Excoffier, L. (2016). Genetic surfing in human populations: from genes to genomes. *Current Opinion in Genetics & Development*, 41, 53-61.

Peischl, S., Dupanloup, I., Kirkpatrick, M., & Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, 22, 5972-5982.

Peischl, S., & Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. *Molecular Ecology*, 24, 2084-2094.

Peischl, S., & Gilbert, K. J. (2020). Evolution of dispersal can rescue populations from expansion load. *American Naturalist*, 195, 349-360.

Peischl, S., Kirkpatrick, M., & Excoffier, L. (2015). Expansion load and the evolutionary dynamics of a species range. *American Naturalist*, 185, E81–E93.

Pennings, P. S., Kryazhimskiy, S., & Wakeley, J. (2014). Loss and recovery of genetic diversity in adapting populations of HIV. *PLoS Genetics*, 10, e1004000.

Perrin, N., & Goudet, J. (2001). Inbreeding, kinship, and the evolution of natal dispersal. In J. Clobert, E. Danchin, A. A. Dhondt and J. D. Nichols (Eds.), *Dispersal*. (pp. 123-142). Oxford: Oxford University Press.

Peter, B. M., & Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution*, 67, 3274-3289.

Petrů, M., Tielbörger, K., Belkin, R., Sternberg, M., & Jeltsch, F. (2006). Life history variation in an annual plant under two opposing environmental constraints along an aridity gradient. *Ecography*, 29, 66-74.

Phillips, B. L. (2009). The evolution of growth rates on an expanding range edge. *Biology Letters*, 5, 802-804.

Phillips, B. L., Brown, G. P., & Shine, R. (2010). Life-history evolution in range-shifting populations. *Ecology*, 91, 1617-1627.

Phillips, B. L., Brown, G. P., Webb, J. K., & Shine, R. (2006). Invasion and the evolution of speed in toads. *Nature*, 439, 803-803.

Picó, F. X., Méndez-Vigo, B., Martínez-Zapater, J. M., & Alonso-Blanco, C. (2008). Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. *Genetics*, 180, 1009-1021.

Prendeville, H. R., Barnard-Kubow, K., Dai, C., Barringer, B. C., & Galloway, L. F. (2013). Clinal variation for only some phenological traits across a species range. *Oecologia*, 173, 421-430.

Pyšek, P. (1997). Compositae as invaders: Better than the others? *Preslia*, 69, 9-22.

Q

QGIS Development Team (2016). *QGIS Geographic Information System*. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R

R Core Team (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197, 573-589.

Rambuda, T. D., & Johnson, S. D. (2004). Breeding systems of invasive alien plants in South Africa: Does Baker's rule apply? *Diversity and Distributions*, 10, 409-416.

Ramos-Onsins, S. E., & Mitchell-Olds, T. (2007). Mlcoalsim: multilocus coalescent simulations. *Evolutionary Bioinformatics*, 3, 41-44.

Raquin, A.-L., Brabant, P., Rhoné, B., Balfourier, F., Leroy, P, & Goldringer, I. (2008). Soft selective sweep near a gene that increases plant height in wheat. *Molecular Ecology*, 17, 741-756.

Rasband, W.S. (1997-2018). ImageJ. National Institutes of Health, Bethesda, Maryland, USA. https://imagej.nih.gov/ij

Rathcke, B., & Lacey, E. P. (1985). Phenological patterns of terrestrial plants. *Annual Review of Ecology, and Systematics*, 16, 179-214.

Riba, M., Mayol, M., Giles, B. E., Ronce, O., Imbert, E., van der Velde, M., … Olivieri, I. (2009). Darwin's wind hypothesis: does it work for plant dispersal in fragmented habitats? *New Phytologist*, 183, 667-677.

Rodríguez-Sánchez, F., Pérez-Barrales, R., Ojeda, F., Vargas, P., & Arroyo, J. (2008). The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews*, 27, 2100-2117.

Ross, M. A., & Harper, J. L. (1972). Occupation of biological space during seedling establishment. *Journal of Ecology*, 60, 77-88.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, DOI: 10.18637/jss.v048.i02.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from *F*-Statistics under isolation by distance. *Genetics*, 145, 1219-1228.

Roux, F., Touzet, P., Cuguen, J., & Le Corre, V. (2006). How to be early flowering: an evolutionary perspective. *Trends in Plant Science*, 11, 375-381.

Roze, D., & Rousset, F. (2005). Inbreeding depression and the evolution of dispersal rates: A multilocus model. *American Naturalist*, 166, 708-721.

Rubio de Casas, R., Kovach, K., Dittmar, E., Barua, D., Barco, B., & Donohue, K. (2012). Seed after-ripening and dormancy determine adult life history independently of germination timing. *New Phytologist*, 194, 868-879.

Ruiz de Clavijo, E. (2001). The role of dimorphic achenes in the biology of the annual weed *Leontodon longirrostris*. *Weed Research*, 41, 275-286.

S

Saikkonen, K., Taulavuori, K., Hyvönen, T., Gundel, P. E., Hamilton, C. E., Vänninen, I., … Helander, M. (2012). Climate change-driven species' range shifts filtered by photoperiodism. *Nature Climate Change*, 2, 239-242.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14, 807-820.

Schaffer, W. M. (1974). Selection for optimal life histories: the effects of age structure. *Ecology*, 55, 391-303.

Schneider, H. E., & Mazer, S. J. (2016). Geographic variation in climate as a proxy for climate change: Forecasting evolutionary trajectories from species differentiation and genetic correlations. *American Journal of Botany*, 103, 140-152.

Schrider, D. R., & Kern, A. D. (2017). Soft sweeps are the dominant mode of adaptation in the human genome. *Molecular Biology and Evolution*, 34, 1863-1877.

Schrider, D. R., Mendes, F. K., Hahn, M. W., & Kern, A. D. (2015). Soft shoulders ahead: spurious signatures of soft and partial selective sweeps result from linked hard sweeps. *Genetics*, 200, 267-284.

Seoane-Zonjic, P., Cañas, R. A., Bautista, R., Gómez-Maldonado, J., Arrillaga, I., Fernández-Pozo, N., … Ávila, C. (2016). Establishing gene models from the Pinus pinaster genome using gene capture and BAC sequencing. *BMC Genomics*, 17, 148.

Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics*, 40, 415-436.

Sexton, J. P., Strauss, S. Y., & Rice, K. J. (2011). Gene flow increases fitness at the warm edge of a species' range. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 11704-11709.

Shaw, A. K., & Kokko, H. (2015). Dispersal evolution in the presence of Allee effects can speed up or slow down invasions. *American Naturalist*, 185, 631-639.

Shimono, Y., & Kudo, G. (2003). Intraspecific variations in seedling emergence and survival of *Potentilla matsumurae* (Rosaceae) between alpine fellfield and snowbed habitats. *Annals of Botany*, 91, 21-29.

Shine, R., Brown, G. P., & Phillips, B. L. (2011). An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 5708-5711.

Siemann, E., & Rogers, W. E. (2001). Genetic differences in growth of an invasive tree species. *Ecology Letters*, 4, 514-518.

Slatkin, M., & Excoffier, L. (2012). Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, 191, 171-181.

Stearns, S. C. (1992). *The Evolution of Life Histories*. Oxford: Oxford University Press.

Stearns, S. C., Ackermann, M., Doebeli, M., & Kaiser, M. (2000). Experimental evolution of aging, growth, and reproduction in fruitflies. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 3309-3313.

Stephens, P. A., Sutherland, W. J., & Freckleton, R. P. (1999). What is the Allee Effect? *Oikos*, 87, 185-190.

Szücs, M., Vahsen, M. L., Melbourne, B. A., Hoover, C., Weiss-Lehman, C., Hufbauer, R. A., & Schoener, T. W. (2017). Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 13501-13506.

T

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105, 437-460.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.

Takou, M., Hämälä, T., Steige, K. A., Koch, E., Dittberner, H., Yant, L., … de Meaux, J. (2019a). Maintenance of adaptive dynamics in a bottlenecked range-edge population that retained out-crossing. *biorxiv*, DOI:10.1101/709873.

Takou, M., Wieters, B., Kopriva, S., Coupland, G., Linstädter, A., & de Meaux, J. (2019b). Linking genes with ecological strategies in *Arabidopsis thaliana*. *Journal of Experimental Botany*, 70, 1141-1151.

Talavera, S., Talavera, M., & Sánchez, C. (2015). Los géneros *Thrincia* Roth y *Leontodon* L. (Compositae, Cichorieae) en Flora Iberica. *Acta Botanica Malacitana*, 40, 344-364.

Tang, D., Ade, J., Frye, C. A., & Innes, R. W. (2005). Regulation of plant defense responses in Arabidopsis by EDR2, a PH and START domain-containing protein. *The Plant Journal*, 44, 245-257.

Thomas, C. D., Bodsworth, E. J., Wilson, R. J., Simmons, A. D., Davies, Z. G., Musche, M., & Conradt, L. (2001). Ecological and evolutionary processes at expanding range margins. *Nature*, 411, 577-581.

Thompson, J. D. (2005). *Plant Evolution in the Mediterranean*. Oxford: Oxford University Press.

Toledo, B., Marcer, A., Méndez-Vigo, B., Alonso-Blanco, C., & Picó, F. X. (2020). An ecological history of the relict genetic lineage of *Arabidopsis thaliana*. *Environmental and Experimental Botany*, 170, 103800.

Tomasini, M., & Peischl, S. (2020). When does gene flow facilitate evolutionary rescue? *Evolution*, 74, 1640-1653.

Travis, J. M. J., Münkemüller, T., Burton, O. J., Best, A., Dytham, C., & Johst, K. (2007). Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and Evolution*, 24, 2334-2343.

Tremetsberger, K., Ortiz, M. A., Terrab, A., Balao, F., Casimiro-Soriguer, R., Talavera, M., & Talavera, S. (2016). Phylogeography above the species level for perennial species in a composite genus. *AoB PLANTS*, 8, plv142.

Tsuchisaka, A., Yu, G., Jin, H., Alonso, J. M., Ecker, J. R., Zhang, X., … Theologis, A. (2009). A combinatorial interplay among the 1-aminocyclopropane-1-carboxylate isoforms regulates ethylene biosynthesis in *Arabidopsis thaliana*. *Genetics*, 183, 979-1003.

V

Vallès, J., Canela, M. A., Garcia, S., Hidalgo, O., Pellicer, J., Sánchez-Jiménez, I., … Garnatje, T. (2013). Genome size variation and evolution in the family Asteraceae. *Caryologia: International Journal of Cytology, Cytosystematics and Cytogenetics*, 66, 221-235.

Van Petegem, K., Moerman, F., Dahirel, M., Fronhofer, E. A., Vandegehuchte, M. L., Van Leeuwen, T., … Bonte, D. (2018). Kin competition accelerates experimental range expansion in an arthropod herbivore. *Ecology Letters*, 21, 225-234.

Venable, D. L. (2007). Bet hedging in a guild of desert annuals. *Ecology*, 88, 1086-1090.

Venable, D. L., & Lawlor, L. (1980). Delayed germination and dispersal in desert annuals: escape in space and time. Oecologia, 46, 272-282.

Verdú, M. & Travesset, A. (2005). Early emergence enhances plant fitness: A phylogenetically controlled meta-analysis. *Ecology*, 86, 1385-1394.

Vidigal, D. S., Marques, A. C. S. S., Willems, L. A. J., Buijs, G., Méndez-Vigo, B., Hilhorst, H. W. M., … Alonso-Blanco, C. (2016). Altitudinal and climatic associations of seed dormancy and flowering traits evidence adaptation of annual life cycle timing in *Arabidopsis thaliana*. *Plant Cell and Environment*, 39, 1737-1748.

Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, 6, 1248-1258.

Villemereuil, P., Gaggiotti, O. E., Mouterde, M., & Till-Bottraud, I. (2016). Common garden experiments in the genomic era: new perspectives and opportunities. Heredity, 116, 249-254.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, 116, 882-892.

Volis, S., Mendlinger, S., & Ward, D. (2002). Adaptive traits of wild barley plants of Mediterranean and desert origin. *Oecologia*, 133, 131-138.

W

Wang, X. J., Hu, Q. J., Guo, X. Y., Wang, K., Ru, D. F., German, D. A., … Liu, J.-Q. (2018). Demographic expansion and genetic load of the halophyte model plant *Eutrema salsugineum*. *Molecular Ecology*, 27, 2943-2955.

Weber, E., & Schmid, B. (1998). Latitudinal population differentiation in two species of *Solidago* (Asteraceae) introduced into Europe. *American Journal of Botany*, 85, 1110-1121.

Weigand, H., & Leese, F. (2018). Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society*, 184, 528-583.

Weiss-Lehman, C., Hufbauer, R. A., & Melbourne, B. A. (2017). Rapid trait evolution drives increased speed and variance in experimental range expansions. *Nature Communications*, 8, 14303.

Wellenreuther, M., & Hansson, B. (2016). Detecting polygenic evolution: problems, pitfalls, and promises. *Trends in Genetics*, 32, 155-164.

Wilczek, A. M., Burghardt, L. T., Cobb, A. R., Cooper, M. D., Welch, S. M., & Schmitt, J. (2010). Genetic and physiological bases for phenological responses to current and predicted climates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 3129-3147.

Willi, Y., Fracassetti, M., Zoller, S., & Van Buskirk, J. (2018). Accumulation of mutational load at the edges of a species range. *Molecular Biology and Evolution*, 35, 781-791.

Willi, Y., Van Buskirk, J. & Hoffmann, A. A. (2006). Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics*, 37,433-58.

Williams, J. L., Kendall, B. E., & Levine, J. M. (2016). Rapid evolution accelerates plant population spread in fragmented experimental landscapes. *Science*, 353, 482-485.

Williamson, M., & Fitter, A. (1996). The varying success of invaders. *Ecology*, 77, 1661-1666.

Wilson, B. A., Petrov, D. A., & Messer, P. W. (2014). Soft selective sweeps in complex demographic scenarios. *Genetics*, 198, 669-684.

Wolf, J. B., & Wade, M. J. (2009). What are maternal effects (and what are they not)? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1107-1115.

Y

Yan, A., & Chen, Z. (2020). The control of seed dormancy and germination by temperature, light and nitrate. *The Botanical Review*, 86, 39-75.

Z

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research, 18, 821-829.

Zhang, M., Zhou, L., Bawa, R., Suren, H., & Holliday, J. A. (2016). Recombination rate variation, hitchhiking, and demographic history shape deleterious load in poplar. *Molecular Biology and Evolution*, 33, 2899-2910.

Zhen, Y., & Ungerer, M. C. (2008). Clinal variation in freezing tolerance among natural accessions of Arabidopsis thaliana. *New Phytologist*, 177, 419-427.