TESI DOCTORAL UPF 2015

Exploring fitness landscapes

Margarita V. Meer

Thesis Advisor
Dr. Fyodor A. Kondrashov

Bioinformatics and Genomics Programme

Centre for Genomic Regulation (CRG)

Barcelona, September 2015

Acknowledgements

There were many people who helped me during this stage of my scientific ontogenesis.

First of all I would like to thank my family for a lot of support and motivation. I would never do it without them believing in me and in "something cool and too hard to understand" I'm doing.

My colleagues at CRG became more than that to me. They've been always there for me when I faced scientific and life challenges. And I have a feeling that they became another family for me here in Barcelona. Inka, Dinara, Peter, Anna, Carla, Dima, Natalia, Masha P., Anna B., Julia, Mateusz, Oriol – I will miss you a lot!

I would also like to thank some people from other labs:
- Maria Tutukina, my biggest supporter, for a lot of her help and for showing an example of a postdoc I want to become
- Silvia Rodriguez for teaching me how to deal with another person's neurons… and other cell types and for being on my side during the fights for the shelves, chairs and pipettes
- Marc Isalan for his suggestions on the experimental design
- Guillaume Filion for our discussions not only of the research projects but also conversations about life in science.
- Dan Tawfik for showing how important it is to believe in yourself and your idea
- Carlo Carolis and his department for being so friendly, helpful and effective in our collaboration

And the most important person in this list is my supervisor Fyodor Kondrashov. It has been more than ten years since I first met him. That day I couldn't imagine myself

working with this great scientist. But here I am. And I am happy that he turned out to be a great person too.

Abstract

Fitness landscape is a concept, which describes the dependence of phenotype on genotype. It was proposed almost a hundred years ago but only recent burst of technologies finally allowed exploring it. We studied different aspects of fitness landscape applying both: computational and experimental approaches. Using mammalian mitochondrial tRNAs we proved that evolution can proceed not only along the ridges of high fitness but also cross the low fitness valleys. Functional analysis of more than 56 000 mutants of green fluorescent protein from *Aequorea victoria* (avGFP) allowed us to describe local fitness landscape around a particular fitness peak. In addition to this we studied a case of population being on a slope – genetic code is undergoing changes in *Methanosarcina* making them stay in sub-optimum.

Keywords:

Fitness landscape, molecular evolution

Resumen

El paisaje adaptativo (fitness landscape) es un concepto que describe la dependencia del fenotipo en el genotipo. Hace mas de cien años que éste concepto fue propuesto, pero es sólo con la reciente expansión de tecnologías que finalmente ha podido ser explorado. Hemos estudiado diferentes aspectos del paisaje adaptativo aplicando tanto procedimientos computacionales como experimentales. Utilizando tRNAs mitocondriales de mamíferos hemos comprobado que la evolución puede proceder no sólo a lo largo de las crestas de elevado fitness sino también a través de los valles con reducido fitness. Análisis funcional de mas de 56000 mutantes de la proteína verde fuorescente de Aequorea Victoria (avGFP) nos permitió describir el paisaje adaptativo local alrededor de un punto máximo específico de fitness. Mas aún, hemos estudiado un caso de una población en una pendiente – el código genético está atravesando cambios en Methanosarcina quedándose en sub-óptimo.

# Contents

1. Introduction

Figure 1: Representation of a fitness landscape. **A**: A fitness landscape can be visualized as a 'mountainous' landscape in three dimensions with genotypes arranged in the x–y plane and fitness on the z axis. The landscape shown is rugged with three fitness peaks separated by fitness 'valleys', and two imaginary evolutionary trajectories are shown by white dots and arrows. **B**: Wright's two-dimensional "field of gene combinations" is shown. Fitness maximum and minimum are represented as "+" and "−", respectively; dotted lines are contours of equal fitness. C: A fitness landscape for a case with four interacting loci. Nodes represent genotypes; 0 and 1 indicate wild-type and mutant amino acids, respectively. Arrows connect genotypes that differ by a single mutation and point towards genotypes with higher resistance. Bold black arrows indicate the 'greedy' walk (which substitutes the existing genotype with the largest-benefit mutation among the mutations available at each step) from wild-type (0000) to the global maximum (1010) [de Visser et al., 2014]

1.1. The concept of fitness landscape

The concept of fitness landscape was proposed by Sewall Wright in 1932 [Wright, 1932]. And since then it remains one of the fundamental concepts in evolutionary biology. The idea is both nice and simple. We can imagine genotype as a point in the space of all genotypes (Fig 1). Usually it is understood as a multidimensional space where every dimension represents some functional unit. And alleles are along the axis. Fitness brings another dimension to this system.

Fitness landscape is basically a map, which in theory can be used to predict the path of evolution. But it is still far not the case since we know too little about it.
First attempts to analyze real fitness landscapes came soon after sequencing. It became possible to compare different variants of sequences and even reconstruct ancestral sequences. Unfortunately it didn't provide much information since the history of navigation of these populations in their fitness landscape remained unknown. So, basically the only conclusion could have been made was that these particular sequences which corresponded to the wild type mostly show examples of fitness peaks.

Most of the studies analysed pairwise mutational interactions or the dependence of mean fitness on the number of mutations.

The way populations can navigate in fitness landscape remained one of the most important evolutionary questions for decades. Obviously evolution can proceed towards some peak of high fitness if there are no obstacles (Fig 2: a, c, d). In this case it can go along the way with constantly increasing fitness (Fig 2: a, d) or pass for some time a neutral region. The question was whether or not it can proceed through a valley of low fitness (Fig 2: b). Will the sub-optimal peak be a dead end for this population? We answer this question in the chapter I of this thesis.
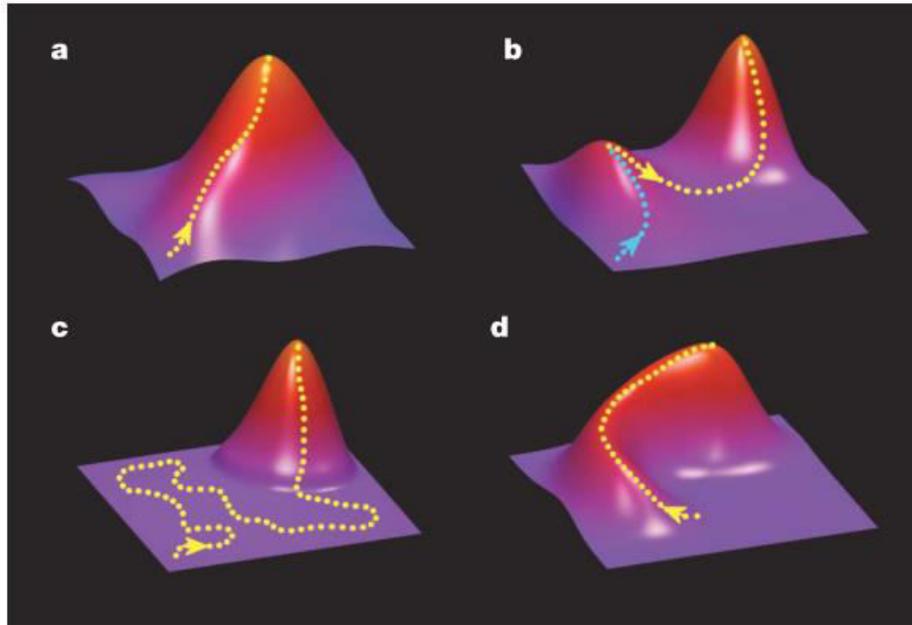
Figure 2. Schematic representations of fitness landscape features.
Fitness is shown as a function of sequence: the dotted lines are mutational paths to higher fitness. a, Single smooth peak. All direct paths to the top are increasing in fitness. b, Rugged landscape with multiple peaks. The yellow path has a fitness decrease that drastically lowers its evolutionary probability. Along the blue path selection leads in the wrong direction to an evolutionary trap. c, Neutral landscape. When neutral mutations are essential, evolutionary probabilities are low. d, Detour landscape. The occurrence of paths where mutations are reverted shows that sequence analysis may fail to show mutations that are essential to the evolutionary history.
[Poelwijk et al., 2007]

1.2. Attempts to investigate fitness landscape

In order to better understand the path evolution can and can not take it was necessarily to perform systematic studies of fitness landscape. There are still just few systematic studies of empirical fitness landscapes. And there are different possible approaches.

The first approach infers qualitative features of the topography, such as its ruggedness, either from patterns of parallel evolution in microbial evolution experiments or from the prevalence of sign epistasis among sets of constructed mutation pairs. This approach allows analyzing large area of genotypic space. Unfortunately, the topographical information it reveals is incomplete and biased by the population dynamic regime used. Another approach involves the systematic analysis of all possible combinations of a

small set of mutations. This approach explores a tiny part of genotypic space around one point, but the information obtained is quite complete and allows to quantify and compare mutational and avolutionary trajectories.

These studies can be also classified based on the sourse of mutations. They can co-occur in extant genotypes or isolated from laboratory evolution experiments. Most of the studies don't estimate real fitness but rather use some proxy of it including growth rate or level of antibiotic resistance.

What information can we get about the topography of real fitness landscapes from these studies? There are two general observations can be made. First, the data sets show a variable, but on average substantial, level of ruggedness. All fitness landscapes are more rugged than expected in the absence of epistasis but less so than expected if mutations had random effects across genetic backgrounds. Second, several data sets show on average diminishing returns epistasis among beneficial mutations, in which benefits are smaller in high fitness backgrounds than in low fitness backgrounds.

Unfortunately no data from previous studies existed for deleterious mutations that are known for their collective negative effect or occurrence in a single gene.

Even though there is still not much data on fitness landscapes some trends can be observed. Mutations with known collective benefit show less ruggedness than mutations for which the combined effect is unknown[Jacquier et al., 2012]. Another remarkable pattern is following: mutations of large effect show greater ruggedness than small effect mutations.The third trend is that intragenic landscapes have greater ruggedness than intergenic landscapes.
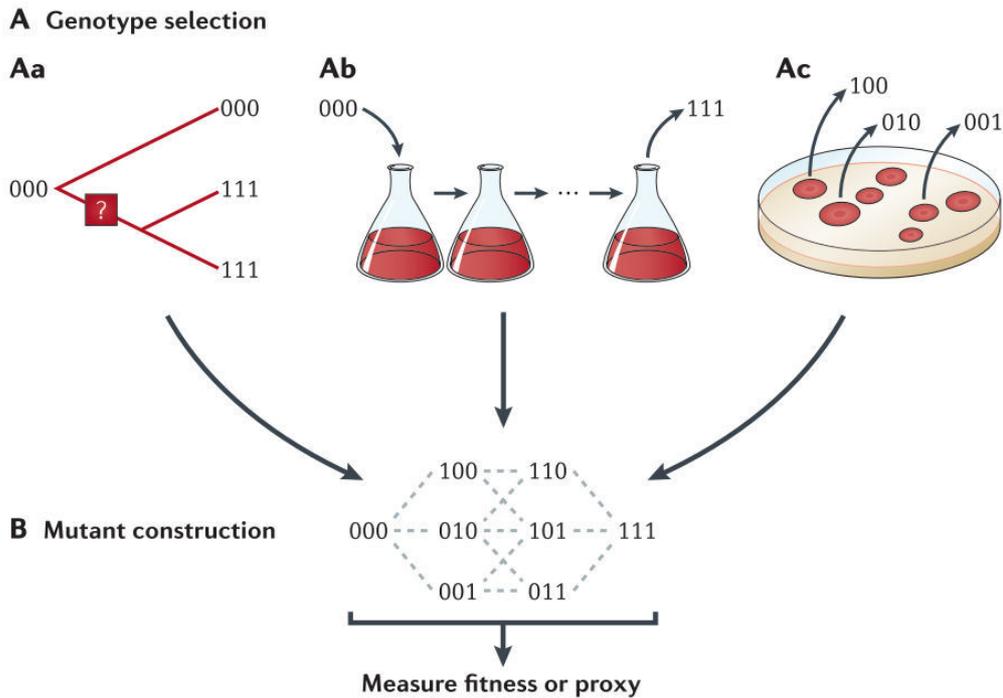
Figure 3. Approaches for the empirical study of fitness landscapes. Experimental approaches for studying small-scale fitness landscapes share three essential components: a set of mutations of interest is identified (part A); mutants are constructed to carry all 2L possible combinations of the L selected mutations (in this case, L = 3, and 0 and 1 indicate the absence and presence of the mutation, respectively) (part B); and the fitness or a fitness proxy (for example, antibiotic resistance) is measured for all genotypes. Mutations of interest can come from three different sources: from phylogenetic analyses that infer the ancestor of extant genotypes (part Aa); from microbial evolution experiments in which mutations co-occur in an evolving lineage (part Ab); or from sets of mutants that each carry a single mutation (part Ac), such as alternative mutations that cause antibiotic resistance. The a posteriori approaches (parts Aa, Ab) are less likely to find much sign epistasis because these mutations have collectively 'survived' the selective pressure, whereas the a priori approach (part Ac) does not suffer from this bias.
[de Visser et al., 2014]

1.3. Experimental approaches

There are two significantly different approaches to experimentally investigate the fitness landscape. It can be either a selection [Palmer et al., 2015] followed by sequencing (Fig 3) or creating a library of mutants and then performing their analysis. The second approach provided opportunity to perform first extensive study of a region of a protein (Fig 4). The method EMPIRIC has been then used to perform analysis of all

single mutants of ubiquitin in the same laboratory and also adopted by other laboratories [Roscoe et al., 2013; Hietpas et al., 2011].
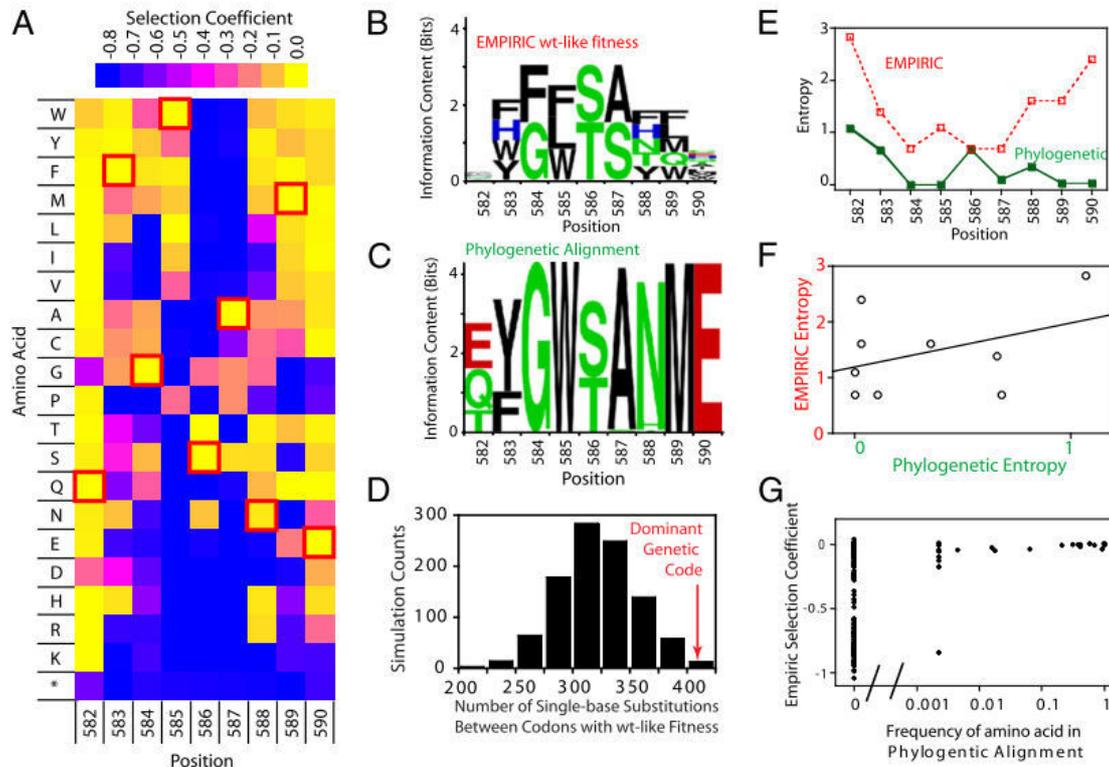

Figure 4. Amino acid profile in phylogenetic alignment poorly predicts EMPIRIC fitness profile. (A) Heat map representation of the EMPIRIC fitness profile with the wild-type amino acids outlined in red. Information content logos generated from amino acids with WT-like EMPIRIC fitness (B) and a phylogenetic alignment of 448 Hsp90 protein sequences (C). (D) The dominant genetic code is optimized for single-base substitutions between codons with WT-like fitness compared with randomly simulated codes (+2.4σ). (E) Distribution of tolerated and phylogenetically observed amino acids expressed as an entropy where zero corresponds to a frozen position and 3 corresponds to unrestrained positions. (F) Relationship between tolerated amino acid profile from EMPIRIC fitness measurements and phylogenetic alignment. Linear regression indicates a very weak correlation with R^2 of 0.15. (G) EMPIRIC fitness analyzed as a function of amino acid prevalence in the phylogenetic alignment. Most amino acids observed in the phylogenetic alignment are well-tolerated when made in the yeast homolog.
[Hietpas et al., 2011]

1.4. Epistasis
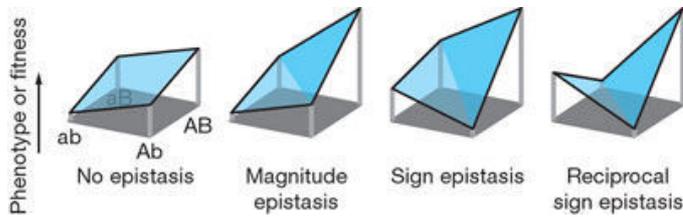
Figure 5
Some variants of epistasis
[Poelwijk et al., 2007]

Nowadays the term epistasis has a plurality of meanings, but was first used by Bateson to describe the masking of the effects of one genetic variant by another [Bateson, 1909]. Later, the term 'epistacy' was used by R.A. Fisher to refer to any statistical deviation from the additive combination of two loci. But then it was replaced with the word epistasis.

This term and the concept behind it play one of the most important roles in evolution. Even though it should be implemented to fitness, usually some measurement of phenotype is considered as a suitable proxy. Than this expected phenotype, such as, for example, growth rate, is compared for a set of mutants.

The simplest case of epistatic interaction between loci can be observed between two diallelic loci (Fig 5). Biological molecules don't provide this type directly. But the closest type of interaction between loci is between corresponding pairs in RNA structures.
The term "genetic interactions" is also often used as a replacement for epistasis if spoken of a situation when different genes are involved.

There are many ways alleles can interact. Thus, epistasis is a very general term covering a lot of possible situations.

Epistatic interactions can be both alleviating (suppressive) or aggravating (increasing).
Interestingly the sign of epistasis can change based on the original effects of the changes making this topic tricky. Interactions can also occur between sequence variants in the same gene ('intramolecular epistasis') or between variants in different genes

('intermolecular epistasis'). Even the presence of gene can be an epistatic issue (Fig 6), as it has been shown for yeast strains by Dowell [Dowell et al., 2010].
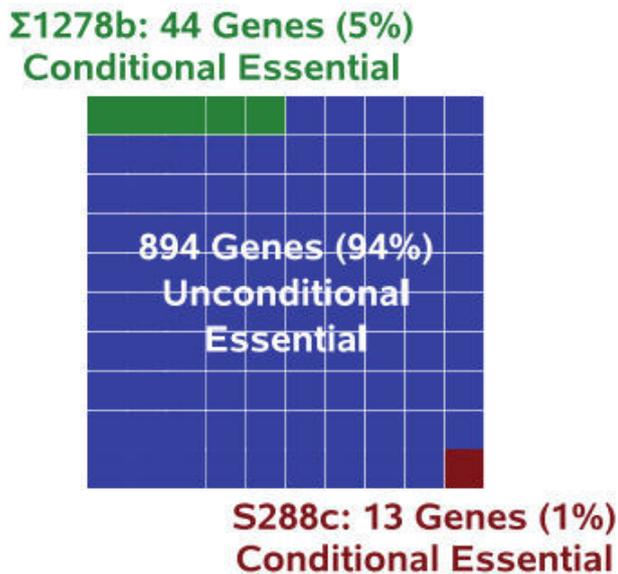

Figure 6.
An example of complex intragenic epistasis:
Most essential genes from the strain S288c of *S. cerevisiae* are also essential in S1278b (94%); however, ~5% are essential only in the S1278b genetic background, whereas ~1% are essential only in S288c.
[Dowell et al., 2010]

The molecular mechanism underlying epistasis is mostly unresolved question. How does it happen that the same mutation turns out to be better in one context but not in another? Is it possible to predict which mutations will show epistatic interactions and how?
Some progress has been made here. Stem structures of RNAs can be an example. Additional clue can be obtained from evolutionary correlated changes.

Individuals of a species typically differ in the sequences of thousands of different proteins, not to mention variation in non-coding regions of the genome. As a result, the consequence of a particular mutation has the potential to be modified by variation in the activities of thousands of different proteins.

One of the evidences for the pervasive nature of epistasis derives from large-scale reverse genetic screens. In this approach pairs of mutations are systematically

combined, and the effects on viability or growth are then determined. The negative and positive interactions between millions of different loss- of-function mutations in yeast have been quantified using this approach. It has been shown that the effects of the mutation can be influenced by perturbing the activity of many additional genes.

Although most genes make few strong negative epistatic interactions, a subset of genes interact with many different loci, and with loci that have many different functions. These genes are called genetic hubs, buffers or capacitors, and are strikingly biased in their molecular functions. They can be chaperones or be members of the chromatin regulation network.

Certainly these genes have the potential to influence the activity of many different proteins, for example in the case of chaperones by assisting protein folding. As an overall trend, there is also a strong correlation between the fitness of a single mutant and the number of epistatic interactions detected for that mutation. The same genes that act as genetic hubs are also important for buffering environmental change and stochastic variation. This way their function of being hubs is a by-product of this function.

2. Overview of the thesis

Appearance of the concept of fitness landscape brought questions on how evolving populations navigate there. The easiest case is when there is no decrease in fitness along the way. It can be either on a flat region under neutral selection or climbing up under positive. But what happens if there is a valley of low fitness on the way to another peak? The question on whether or not evolution can proceed through such a valley remained one of the long-standing questions in evolution. To answer it we decided to study the simplest case – when there are just two loci. An example of this can be observed in stems of RNAs. Corresponding nucleotides can form a Watson-Crick (WC) pair (G-C or A-U) or decrease stability of the stem structure. So we decided to study evolution in these pairs focusing on stems of mitochondrial tRNAs of mammals. This allowed us to show that evolution can proceed through the valleys of low fitness indeed. The comparison of data on evolution and polymorphisms allowed us to estimate the depth of these valleys. This work is presented in the chapter XX of this thesis.

In the next project we decided to focus on the fitness landscape of a particular molecule, since it can be directly linked the fitness of the organism if the functionality of this molecule is essential.

Even though we have a lot of data on sequence variants now usually they correspond to distant functional sequences. Thus we can observe not just a small fraction of a fitness landscape of a molecule, but a set of distant points without having data on what is between them. This way we loose a lot of important information including the general ruggedness of the fitness landscape, the angle of the slope of a particular fitness peak, amount of epistasis. Recently there were several attempts to perform a robust study of fitness landscapes. Some of them investigated a very small space analyzing only single mutants. Another ones focused on just a region of a gene or on a set of functionally or evolutionary important sites. In our project we performed functional analysis of more

than 56 000 mutants of green fluorescent protein from *Aequorea victoria* (avGFP) including variants up to ten changes from the original wild type sequence. This approach allowed us to explore relatively big space in the fitness landscape of avGFP. This project is presented in the chapter XX.

The clade of Methanosarcina provides us with an interesting case of genetic code undergoing changes. In this case TAG/UAG has ambiguous function – it is for both termination of translation and as a coding codon for amino acid pyrrolysine. There is no evidence for existence of the regulatory mechanism. Thus appearance of TAG codon in a particular coding sequence brings the fitness of it down but not completely. During every round of translation UAG can be recognized as a stop-codon or as a coding codon. Thus some fraction of proteins synthetized from this mRNA will be too short/long but some will have appropriate length. The study of evolution of this codon is presented in the chapter XX.

Chapters I and II in the Appendix present papers resulted from collaborations. In the first one we present a case of mistaken identification of bird species Acrocephalus orinus. I participated in the experimental procedures. In the second one we study the origin and evolution of a new biochemical function - Methacrylate Redox System.

3. Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness

4. Local fitness landscape of the green fluorescent protein

Karen S. Sarkisyan[1,2,3]*, Dmitry A. Bolotin[1,3]*, Margarita V. Meer[4,5], Dinara R. Usmanova[4,5,6], Alexander S. Mishin[1,2], George V. Sharonov[1,7], Dmitry N. Ivankov[4,5,8], Nina G. Bozhanova[1], Mikhail S. Baranov[1], Onuralp Soylemez[4,5], Natalya S. Bogatyreva[4,5,8], Peter K. Vlasov[4,5], Evgeny S. Egorov[1], Maria D. Logacheva[9], Alexey S. Kondrashov[9,10], Dmitry M. Chudakov[1,3], Ekaterina V. Putintseva[1,3], Ilgar Z. Mamedov[1,3], Dan S. Tawfik[11], Konstantin A. Lukyanov[1,2], Fyodor A. Kondrashov[4,5,12]

[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Miklukho-Maklaya 16/10, 117997 Moscow, Russia*

[2] *Nizhny Novgorod State Medical Academy, Minin Sq. 10/1, 603005 Nizhny Novgorod, Russia*

[3] *Central European Institute of Technology, Masaryk University, Brno, Czech republic.*

[4] *Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG) 88 Dr. Aiguader, 08003 Barcelona, Spain.*

[5] *Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain.*

[6] *Moscow Institute of Physics and Technology, Institutskiy pereulok 9, g.Dolgoprudny, 141700, Russia*

[7] *Faculty of Medicine, Lomonosov Moscow State University, Lomonosov ave. 31/5 119192 Moscow, Russia*

[8] *Laboratory of Protein Physics, Institute of Protein Research of the Russian Academy of Sciences, 4 Institutskaya str., Pushchino, Moscow region, 142290, Russia*

[9] *Department of Bioinformatics and Bioengineering, Moscow State University, Moscow, Russia*

[10] *Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI*

[11] *Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel.*

[12] *Institució Catalana de Recerca i Estudis Avançats (ICREA), 23 Pg. Lluís Companys, 08010 Barcelona, Spain.*

(Under second revision in Nature)

Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, Bogatyreva NS, Vlasov PK, Egorov ES, Logacheva MD, Kondrashov AS, Chudakov DM, Putintseva EV, Mamedov IZ, Tawfik DS, Lukyanov KA, Kondrashov FA. Local fitness landscape of the green fluorescent protein. Nature. 2016 May 19;533(7603):397-401. doi:10.1038/nature17995.

5. Evolution of TAG codon in Methanosarcina

Margarita Meer[a], Inna Povolotskaya[a], Peter Vlasov[a], Viktoria Shcherbakova[b],

Fyodor Kondrashov[a]

a. Centre for Genomic Regulation (CRG) and Universitat Pompeu Fabra (UPF), Barcelona, Spain

b. Skryabin Instiute of Biochemistry and Physiology of Microorganisms, Pushchino, Russia

margarita.meer@crg.eu

(Paper in preparation)

Abstract

TAG/UAG codon is ambifunctional in the clade of *Methanosarcina* being used as both: translation terminator and a coding codon. No system of *in vivo* distinguishing between these two functions has been reported. Thus, current state of TAG codon provides us with an example of genetic code being changed. We have performed a study of evolution of this codon showing that this is a case of mutational-selection balance with TAG codons being strongly disfavored.

Keywords:

Introduction

Pyrrolysine (Pyl, O) is the 22nd amino acid known to be included into proteins during translation [Hao et al., 2002]. It can be found mostly in methanogenic archaea of the genus Methanosarcina. Pyrrolysine is encoded by TAG codon [Ibba et al., 2002]. At the same time as it has been shown experimentaly TAG still retains the stop codon function in this genus [Alkalaeva et al., 2009].

In contrast to the Selenocysteine (Sec) coding system, where the SECIS structure determines the Sec-coding function of a specific TGA codon, such system has not been found for Pyl-coding TAG codons [Namya et al., 2007]. There are also other differences between Pyl and Sec, including the way they are synthetized. Whereas Sec is a product of modification of the amino acid Serine, which is already attached to Sec-tRNA, Pyl has a pathway as the canonical 20 amino acids. First it is synthetized and only then loaded to the corresponding tRNA by Pyrrolysyl-tRNA synthetase [Krzycki, 2005].

Selenocysteine needs Selenocysteine-specific elongation factor to be incorporated into the polypeptide chain [Forchhammer, 1989]. Pyl doesn't need something like this [Namya et al., 2007]. Thus, in some sense Pyl behaves more similar to the canonical amino acids.
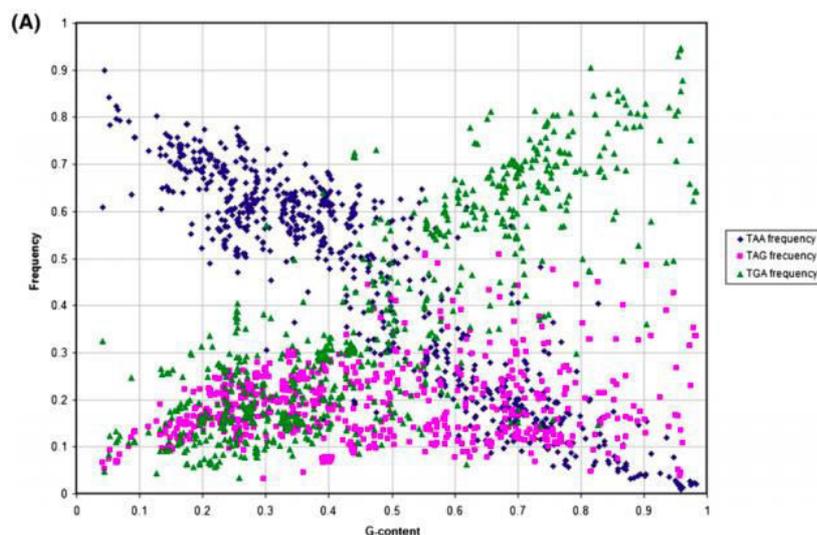

Figure 1. Dependence of stop codon usage on GC-content. TAA is shown is blue, TAG in magenta and TGA in green.

[Povolotskaya et al., 2012]

TAG/UAG codon in Methanosarcina is bi-functional with the Pyl-coding and stop codon functions being in a competitive inhibition-like state [Lobanov et al., 2010]. Thereby from a single mRNA at every read through, a TAG/UAG codon has a chance of being recognized by translation machinery as a stop codon or as a Pyl codon.

Most of prokaryotic genomes contain around 20% of genes terminated by TAG (Fig 1). But for Methanosarcina this amount is reduced to just a few percents [Zhang et al., 2005]. This perfectly fits with such an inefficient system, suggesting TAG codons to be deleterious.

Here we provide analysis of the evolution of TAG codons in this genus.

Results and discussion

The amount of TAG containing genes in Pyl-coding Achaea from thee clade of *Methanosarcina* is much lower than in other prokaryotic species signifying that there has been a decrease of their fraction in the past. But now there are three possibilities. In the first one TAG codon is still highly deleterious as it should be for a dual functional and not regulated system. Another opportunity is that our knowledge of the Pyl incorporation is not good enough, and TAG became a pure Pyl codon. In this case the fraction of TAG should be increasing. And under the third scenario the amount of TAG should be stable: these variants arise due to natural mutagenesis but they are eliminated with the speed similar to the rate of emergence.

Most of these changes occur in the end of coding sequences – in a region known to be less conservative than the upstream region. This brings a complication to genomic comparisons since different species turn out to be too distant to trace small changes in the end of the sequence. A solution can be to use different strains. But in this case there can be too few changes.

To study the evolution of TAG codon we used 140 genomes of *Methanosarcina*

sp. from the Genbank and sequenced 14 more genomes. We divided them into three groups based on their similarity (Fig 2,3,4). *M.mazei* turned out to be species with the biggest amount of genomes available. We had 131 strains. But these strains were very similar to each other. Some of them are rather different samples of the same strain. We managed to obtain just 11 *M.barkeri* genomes. But they were on average more distant.
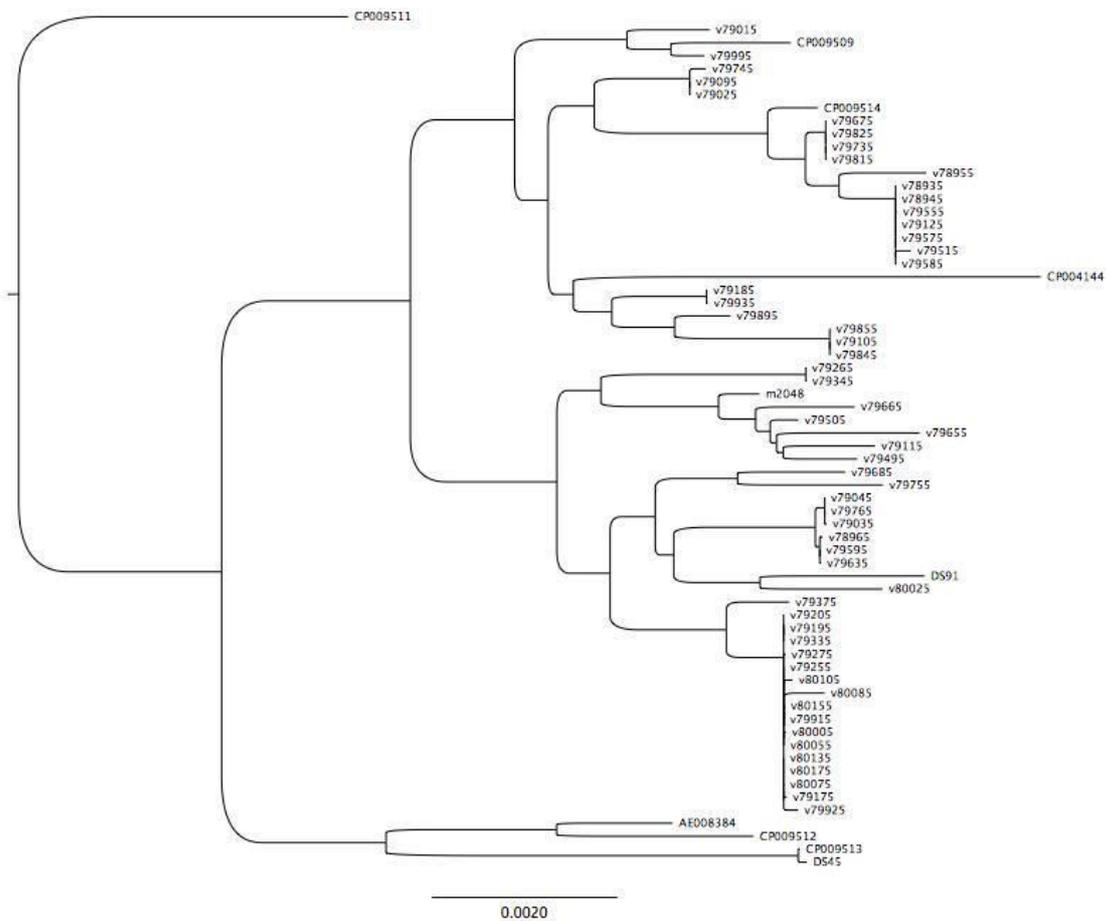
Figure 2: Phylogeny of *M.mazei* strains

Other 12 genomes were even more distant. This group included several strains of *M. thermophila*, *M. acetivorans* and some other strains including the ones of undefined species.

We have annotated de-novo sequenced genomes and re-annotated the ones existing in the GenBank. We considered TAG as a coding codon. This way none of coding sequence terminated with it.

The amount of genes with TAG codons turned out to vary a bit between the

three groups (Tab 1).

Table 1. Fraction of genes with following codons

	M.mazei	M.barkeri	M.sp
TAG	2,2	1,7	1,2
TAA	56,8	59,6	50,5
TGA	43,2	40,4	49,5
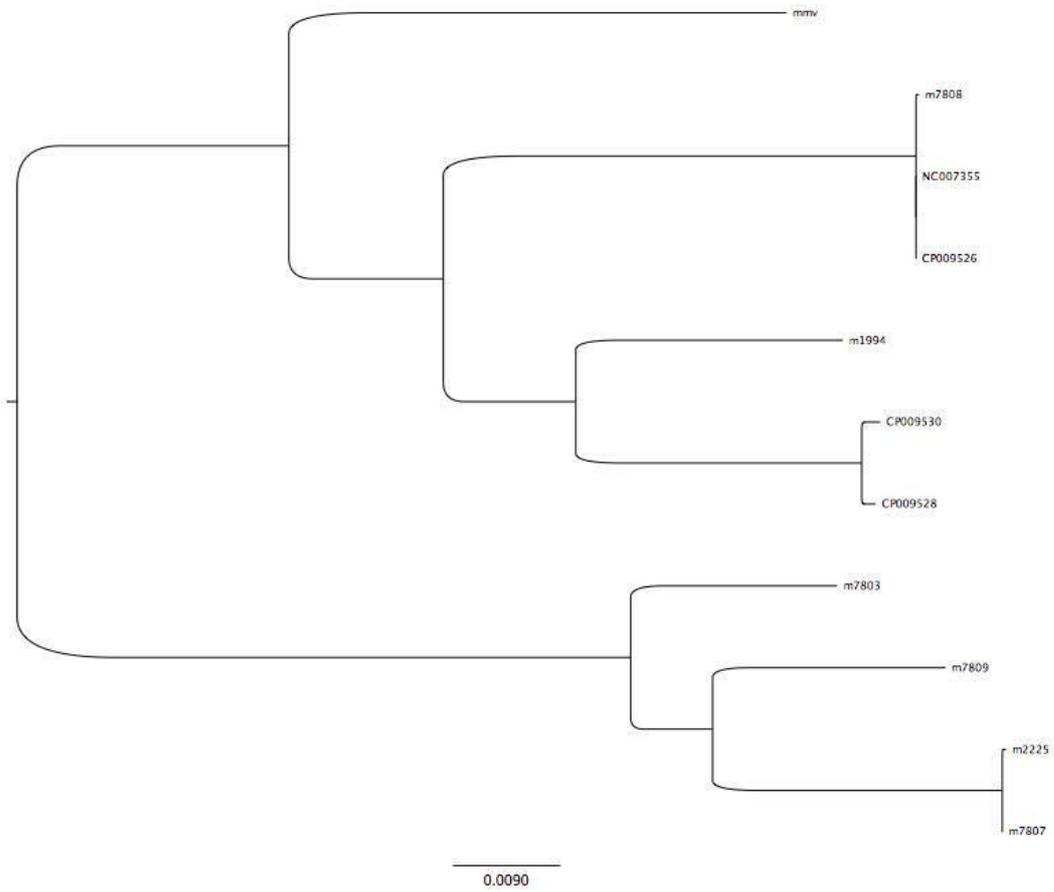
Figure 3: Phylogeny of *M.barkeri* strains
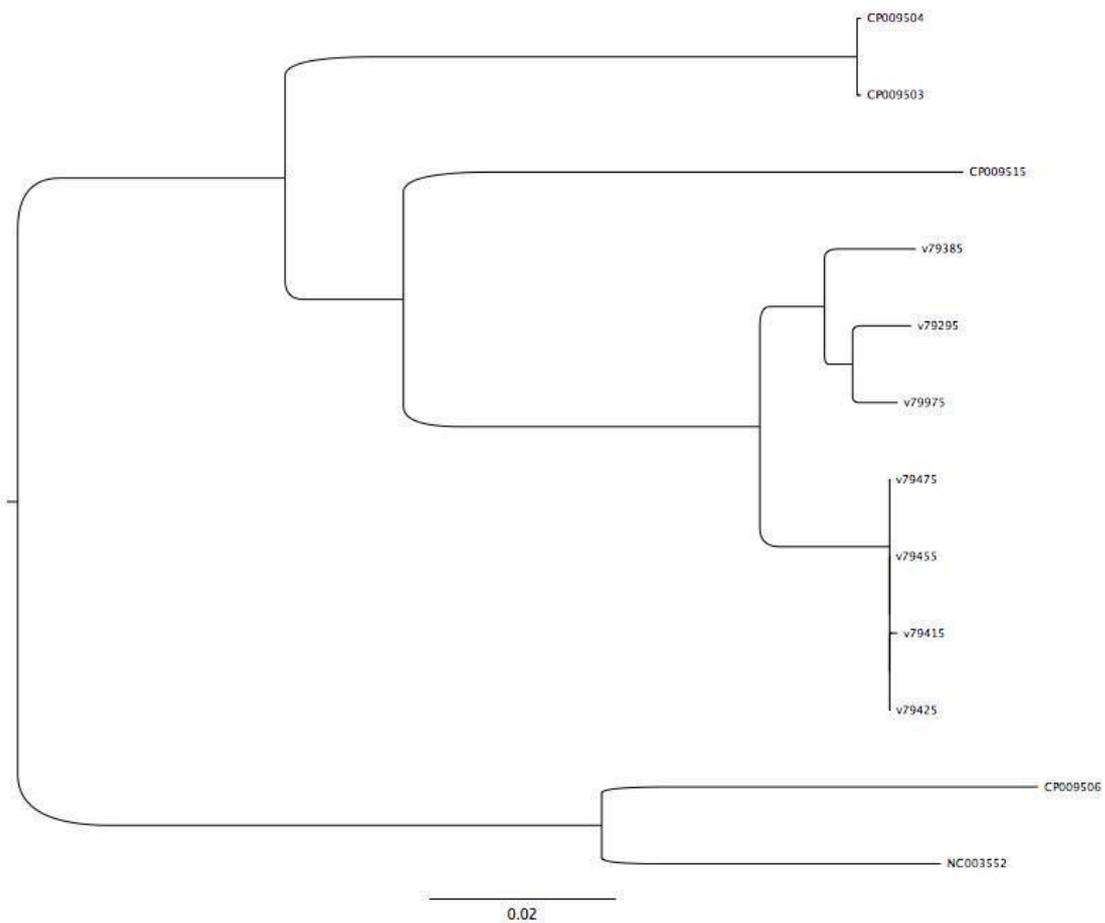
Figure 4: Phylogeny of *Methanosarcina sp.* strains

To check whether or not the amount of TAG codon is changing in these genomes, we performed a phylogenetic analysis. *In M.barkeri* 55 TAG were gained and 52 lost, meanwhile in other two groups the situation was different: *M. mazei* had 59 gains versus 28 loses. The situation in *M.sp* group was similar to *M.mazei*: in 62 gains and 28 loses. Does this mean that TAG stopped being disfavored?

To answer this question we decided to measure the rate of evolution. Since changes in TAG are caused often not just by point mutations, we considered also all types of changes for other codons. The rate of changes of TAA and TGA was similar to that for coding codons. Whereas the rate of evolution of TAG was increased 7.8, 6.7 and 5.3 times for *M.mazei*, *M.barkeri* and *M.sp.* correspondingly.

Thus we observe a situation when the amount of TAG codon increases even

though it is strongly deleterious.

Table 2. Rate of changes of different codons across the coding sequences of the genomes

Codon	M.mazei	M.barkeri	M.sp
AAA	0,10	2,34	2,64
AAC	0,15	4,73	4,09
AAG	0,15	4,16	4,35
AAT	0,17	3,86	3,89
ACA	0,15	3,43	4,52
ACC	0,16	5,41	5,15
ACG	0,25	6,11	6,36
ACT	0,18	3,95	5,05
AGA	0,16	3,71	4,66
AGC	0,13	4,77	4,58
AGG	0,14	4,84	4,77
AGT	0,18	3,92	4,49
ATA	0,12	2,90	3,76
ATC	0,15	4,84	4,67
ATG	0,06	1,21	1,38
ATT	0,14	3,07	3,79
CAA	0,24	3,28	3,82
CAC	0,16	4,52	3,97
CAG	0,07	2,39	2,34
CAT	0,16	3,34	3,85
CCA	0,25	4,40	5,56
CCC	0,20	5,68	5,24
CCG	0,21	6,19	5,78
CCT	0,14	3,00	4,35
CGA	0,25	5,04	5,86
CGC	0,14	4,86	4,50
CGG	0,21	6,58	6,35
CGT	0,21	4,04	4,50
CTA	0,42	6,94	7,95
CTC	0,15	4,72	4,51
CTG	0,12	4,54	4,29
CTT	0,10	2,32	3,15
GAA	0,11	2,40	2,71
GAC	0,16	4,91	4,61

GAG	0,15	4,62	4,67
GAT	0,17	3,63	4,04
GCA	0,11	2,94	3,70
GCC	0,17	5,43	5,28
GCG	0,27	7,44	7,06
GCT	0,14	3,75	4,65
GGA	0,14	3,09	4,08
GGC	0,17	5,49	5,64
GGG	0,17	5,62	5,27
GGT	0,21	4,80	5,62
GTA	0,17	3,75	4,65
GTC	0,17	5,39	5,53
GTG	0,17	4,61	4,61
GTT	0,15	3,58	4,27
TAA	0,16	2,61	2,55
TAC	0,13	3,58	3,43
TAG	1,25	27,15	22,94
TAT	0,12	2,36	2,70
TCA	0,14	3,02	4,13
TCC	0,14	4,68	4,30
TCG	0,20	5,40	5,54
TCT	0,15	3,38	4,44
TGA	0,14	2,77	2,78
TGC	0,11	3,52	3,47
TGG	0,02	0,26	0,34
TGT	0,16	2,91	3,38
TTA	0,14	3,34	3,52
TTC	0,11	3,19	2,99
TTG	0,24	4,63	4,72
TTT	0,09	1,93	2,32

Materials and methods

140 genomes of *Methanosarcina sp.* were downloaded from the GenBank

Cultures and DNA of 14 *Methanosarcina* strains were obtained from DSMZ (https://www.dsmz.de/).

These of the genomes were sequenced and assembled de novo. The genomes were sequenced using Illumina HiSeq technology. A combination of approaches and programs was used for the assembly.

SOAPdenovo [Luo et al.: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 2012 1:18.], velvet [Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly usingde Bruijn graphs. Genome Res. 2008 May;18(5):821-9.] and SPAdes [Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A,
Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol. 2013 Oct;20(10):714-37.] were used as assemblers.

GapCloser was used to improve the assemblies.

Open reading frames were predicted using Glimmer [S. Salzberg, A. Delcher, S. Kasif, and O. White. Microbial gene identification using interpolated Markov models, Nucleic Acids Research 26:2 (1998), 544-548.] with only TAA and TGA being set as stop codons.

Genes were clustered to orthological goups based on the results of reciprocal best blast hit approach [Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410].

Sequences were aligned with Muscle [Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 19;32(5):1792-7.]. TAA, TGA and TAG codons were pseudotranslated as amino acid X.

Phylogenetic trees were reconstructed using PhyML ["New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0."Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. Systematic Biology, 59(3):307-21, 2010].

Ancestry reconstruction and trace of evolutionary changes were performed with modified PAML package [Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol., 2007 Aug;24(8):1586-91. Epub 2007 May 4.]. TAA, TGA and TAG were set as coding codons and they were included into analysis.

All aligntments with genes containing TAG were manually checked. All changes from/to TAG codon were also manually verified.

Data handling and analysis was done with scripts written in Perl.

6. Discussion

There are different aspects of fitness landscape, which can be issued. For decades it was possible to investigate it just theoretically. Luckily we live in the era when we can easily collect or even create a lot of data to perform extensive studies.

Proposing the concept of fitness landscape Sewall Wright greatly underestimated the number of loci [Wright, 1932]. "10 allelomorphs in 1000 loci" he mentions in his paper is an underestimation not even for an organism but for a single protein! There is a tendency to study fitness landscapes of particular proteins or RNAs. But one should always take into account the existence of intragenic epistasis. So, the fitness landscape obtained for an isolated or studied in an artificial system protein might differ from the native one. Taking a protein, which doesn't interact with many other biomolecules might be a solution. This is also one of the reasons why we took GFP.

We had opportunity to look at the epistasis for two types of molecules – a protein and a set of tRNAs. And it is not surprising that we observed more epistasis in the stem structures of tRNAs than studying the local fitness of GFP since stems are very special epistatisis-prone structures. The molecular mechanism of epistasis in the stems of RNAs is well understood and it is based the stability of the structure. Much less in understood about the proteins and about loops of RNAs. A project similar to the one we did with GFP has recently been done on a tRNA, significantly increasing the amount of information about epistasis and evolution in RNAs [Chuan Li, unpublished]. Publications on extensive and systematic investigation of fitness landscapes come out so often these days that it makes me being optimistic about us being able to understand and start using it soon.

The ability to predict phenotype from the genotype is a philosopher's stone of modern geneticists. Checking effect of every single mutation is challenging. But even this turns

out to be not enough due to epistasis. But how worried we really should be about it? How much does it influence evolution? In this thesis we tried to provide some answers highlighting areas in fitness landscapes but there is still much to be done to really shed the light.

TAG codon in *Methanosarcina* is an example of how the change of genetic code can proceed. We see that genomes undergoing this switch decreased the frequency of this codon and it is still under strong negative selection. Leading to a fraction of proteins synthetized from a particular mRNA being inappropriate length it reduces the fitness of these proteins in a non-trivial manner. A mutation in a coding sequence influences the expression level.

7. Conclusions

1. More than a half of changes in stems of mammalian mitochondrion tRNAs are involved in compensatory changes with AT<->GC being the most common ones

2. The path of changes in a compensatory evolution of mammalian mt-tRNAs stem structures can be explained by Shifting-Balance model but not by Dobzhansky-Muller

3. Fitness peak of GFP is narrow with elimination of fluorescence after just five random mutations

4. Epistasis influenced 30% of variants of GFP mutants we have analysed

5. 10% of single mutations kill the function of GFP, but at least 10% of them are examples of CPDs

6. TAG codon is strongly disfavored in Methanosarcina however it is gained overall

8. Bibliography

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. a C., Andersson, M., … Zink, A. (2011a). Inclusive fitness theory and eusociality. *Nature*, *471*(7339), E1–E4; author reply E9–E10. doi:10.1038/nature09831

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. a C., Andersson, M., … Zink, A. (2011b). Inclusive fitness theory and eusociality. *Nature*, *471*(7339), E1–E4; author reply E9–E10. doi:10.1038/nature09831

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. *Genetics*, *139*(2), 1067–1076.

Amini, S., Hottes, A. K., Smith, L. E., & Tavazoie, S. (2011a). Fitness landscape of antibiotic tolerance in pseudomonas aeruginosa biofilms. *PLoS Pathogens*, *7*(10). doi:10.1371/journal.ppat.1002298

Amini, S., Hottes, A. K., Smith, L. E., & Tavazoie, S. (2011b). Fitness landscape of antibiotic tolerance in pseudomonas aeruginosa biofilms. *PLoS Pathogens*, *7*(10). doi:10.1371/journal.ppat.1002298

Assis, R. (2014a). Strong Epistatic Selection on the RNA Secondary Structure of HIV. *PLoS Pathogens*, *10*(9), e1004363. doi:10.1371/journal.ppat.1004363

Assis, R. (2014b). Strong Epistatic Selection on the RNA Secondary Structure of HIV. *PLoS Pathogens*, *10*(9), e1004363. doi:10.1371/journal.ppat.1004363

Athavale, S. S., Spicer, B., & Chen, I. a. (2014a). Experimental fitness landscapes to understand the molecular evolution of RNA-based life. *Current Opinion in Chemical Biology*, *22C*, 35–39. doi:10.1016/j.cbpa.2014.09.008

Athavale, S. S., Spicer, B., & Chen, I. a. (2014b). Experimental fitness landscapes to understand the molecular evolution of RNA-based life. *Current Opinion in Chemical Biology*, *22C*, 35–39. doi:10.1016/j.cbpa.2014.09.008

Atkins, J. F., & Gesteland, R. (2002). Biochemistry. The 22nd amino acid. *Science (New York, N.Y.)*, *296*(5572), 1409–1410. doi:10.1126/science.1073339

Bateson W. 1909. Mendel's principles of heredity. Cambridge (UK): Cambridge University Press. Available

Bank, C., Hietpas, R. T., Jensen, J. D., & Bolon, D. N. a. (2014a). A Systematic Survey of an Intragenic Epistatic Landscape. *Molecular Biology and Evolution*, *32*(1), 229–238. doi:10.1093/molbev/msu301

Bank, C., Hietpas, R. T., Jensen, J. D., & Bolon, D. N. a. (2014b). A Systematic Survey of an Intragenic Epistatic Landscape. *Molecular Biology and Evolution*, *32*(1), 229–238. doi:10.1093/molbev/msu301

Bank, C., Hietpas, R. T., Jensen, J. D., & Bolon, D. N. a. (2014c). A Systematic Survey of an Intragenic Epistatic Landscape. *Molecular Biology and Evolution*, *32*(1), 229–238. doi:10.1093/molbev/msu301

Bank, C., Hietpas, R. T., Jensen, J. D., & Bolon, D. N. a. (2014d). A Systematic Survey of an Intragenic Epistatic Landscape. *Molecular Biology and Evolution*, *32*(1), 229–238. doi:10.1093/molbev/msu301

Bashey, F., & Bashey, F. (2015a). Within-host competitive interactions as a mechanism for the maintenance of parasite diversity.

Bashey, F., & Bashey, F. (2015b). Within-host competitive interactions as a mechanism for the maintenance of parasite diversity.

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., & Tawfik, D. S. (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature*, *444*(7121), 929–932. doi:10.1038/nature05385

Bitbol, A.-F., & Schwab, D. J. (2014a). Quantifying the Role of Population Subdivision in Evolution on Rugged Fitness Landscapes. *PLoS Computational Biology*, *10*(8), e1003778. doi:10.1371/journal.pcbi.1003778

Bitbol, A.-F., & Schwab, D. J. (2014b). Quantifying the Role of Population Subdivision in Evolution on Rugged Fitness Landscapes. *PLoS Computational Biology*, *10*(8), e1003778. doi:10.1371/journal.pcbi.1003778

Borrel, G., Gaci, N., Peyret, P., O'Toole, P. W., Gribaldo, S., & Brugère, J. F. (2014). Unique characteristics of the pyrrolysine system in the 7th order of methanogens: Implications for the evolution of a genetic code expansion cassette. *Archaea*, *2014*. doi:10.1155/2014/374146

Bridgham, J. T., Ortlund, E. a, & Thornton, J. W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*, *461*(7263), 515–519. doi:10.1038/nature08249

Chen, Y., & Stephan, W. (2003a). Compensatory evolution of a precursor messenger RNA secondary structure in the Drosophila melanogaster Adh gene. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11499–11504. doi:10.1073/pnas.1932834100

Chen, Y., & Stephan, W. (2003b). Compensatory evolution of a precursor messenger RNA secondary structure in the Drosophila melanogaster Adh gene. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(20), 11499–11504. doi:10.1073/pnas.1932834100

Chilingarian, a., Hovsepyan, G., Khanikyanc, G., Reymers, a., & Soghomonyan, S. (2015a). Lightning origination and thunderstorm ground enhancements terminated by the lightning flash. *EPL (Europhysics Letters)*, *110*(4), 49001. doi:10.1209/0295-5075/110/49001

Chilingarian, a., Hovsepyan, G., Khanikyanc, G., Reymers, a., & Soghomonyan, S. (2015b). Lightning origination and thunderstorm ground enhancements terminated by the lightning flash. *EPL (Europhysics Letters)*, *110*(4), 49001. doi:10.1209/0295-5075/110/49001

Cobucci-Ponzano, B., Rossi, M., & Moracci, M. (2012). Translational recoding in archaea. *Extremophiles*, *16*(6), 793–803. doi:10.1007/s00792-012-0482-8

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., … Boone, C. (2010a). The genetic landscape of a cell. *Science (New York, N.Y.)*, *327*(5964), 425–431. doi:10.1126/science.1180823

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., … Boone, C. (2010b). The genetic landscape of a cell. *Science (New York, N.Y.)*, *327*(5964), 425–431. doi:10.1126/science.1180823

Crona, K., Greene, D., & Barlow, M. (2014a). NIH Public Access, 1–10. doi:10.1016/j.jtbi.2012.09.028.THE

Crona, K., Greene, D., & Barlow, M. (2014b). NIH Public Access, 1–10. doi:10.1016/j.jtbi.2012.09.028.THE

Crow, J. F. (2010). On epistasis: why it is unimportant in polygenic directional selection. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1544), 1241–1244. doi:10.1098/rstb.2009.0275

Crow, J. F., & Kimura, M. (1979). Efficiency of truncation selection. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(1), 396–399. doi:10.1073/pnas.76.1.396

Cuevas, M., Elena, S. F., Hillung, J., & Elena, S. F. (2015a). Evaluating the within-host fitness effects of mutations fixed during virus adaptation to different ecotypes of a new host.

Cuevas, M., Elena, S. F., Hillung, J., & Elena, S. F. (2015b). Evaluating the within-host fitness effects of mutations fixed during virus adaptation to different ecotypes of a new host.

De Visser, J. a. G. M., Cooper, T. F., & Elena, S. F. (2011). The causes of epistasis. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1725), 3617–3624. doi:10.1098/rspb.2011.1537

De Visser, J. A. G. M., & Krug, J. (2014a). Empirical fitness landscapes and the predictability of evolution. *Nature Reviews. Genetics*, *15*(7), 480–90. doi:10.1038/nrg3744

De Visser, J. A. G. M., & Krug, J. (2014b). Empirical fitness landscapes and the predictability of evolution. *Nature Reviews. Genetics*, *15*(7), 480–490. doi:10.1038/nrg3744

Dean, A. M., & Thornton, J. W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. *Nature Reviews. Genetics*, *8*(9), 675–688. doi:10.1038/nrg2160

DePristo, M. a, Weinreich, D. M., & Hartl, D. L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Reviews. Genetics*, *6*(9), 678–687. doi:10.1038/nrg1672

Dowell, R. D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., … Boone, C. (2010a). Genotype to phenotype: a complex problem. *Science (New York, N.Y.)*, *328*(5977), 469. doi:10.1126/science.1189015

Dowell, R. D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., … Boone, C. (2010b). Genotype to phenotype: a complex problem. *Science (New York, N.Y.)*, *328*(5977), 469. doi:10.1126/science.1189015

Firnberg, E., Labonte, J. W., Gray, J. J., & Ostermeier, M. (2014a). A comprehensive, high-resolution map of a Gene's fitness landscape. *Molecular Biology and Evolution*, *31*(6), 1581–1592. doi:10.1093/molbev/msu081

Firnberg, E., Labonte, J. W., Gray, J. J., & Ostermeier, M. (2014b). A comprehensive, high-resolution map of a Gene's fitness landscape. *Molecular Biology and Evolution*, *31*(6), 1581–1592. doi:10.1093/molbev/msu081

Firnberg, E., Labonte, J. W., Gray, J. J., & Ostermeier, M. (2014c). A comprehensive, high-resolution map of a Gene's fitness landscape. *Molecular Biology and Evolution*, *31*(6), 1581–1592. doi:10.1093/molbev/msu081

Forchhammer, K., Leinfelder, W., & Böck, a. (1989). Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. *Nature*, *342*(6248), 453–456. doi:10.1038/342453a0

Gaston, M. a., Jiang, R., & Krzycki, J. a. (2011). Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Current Opinion in Microbiology*, *14*(3), 342–349. doi:10.1016/j.mib.2011.04.001

Hansis, M., & Heilmann, S. (2010). What can we learn from patients? *Der Unfallchirurg*, *113*(3), 247–248. doi:10.3109/02813438509017724

Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. a, & Chan, M. K. (2002a). A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science (New York, N.Y.)*, *296*(5572), 1462–1466. doi:10.1126/science.1069556

Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. a, & Chan, M. K. (2002b). A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science (New York, N.Y.)*, *296*(5572), 1462–1466. doi:10.1126/science.1069556

Hietpas, R. T., Jensen, J. D., & Bolon, D. N. a. (2011a). Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7896–7901. doi:10.1073/pnas.1016024108

Hietpas, R. T., Jensen, J. D., & Bolon, D. N. a. (2011b). Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7896–7901. doi:10.1073/pnas.1016024108

Hinkley, T., Martins, J., Chappey, C., Haddad, M., Stawiski, E., Whitcomb, J. M., … Bonhoeffer, S. (2011a). A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nature Genetics*, *43*(5), 487–489. doi:10.1038/ng.795

Hinkley, T., Martins, J., Chappey, C., Haddad, M., Stawiski, E., Whitcomb, J. M., … Bonhoeffer, S. (2011b). A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nature Genetics*, *43*(5), 487–489. doi:10.1038/ng.795

Ibba, M., & Söll, D. (2002). Genetic code: Introducing pyrrolysine. *Current Biology*, *12*(13), 464–466. doi:10.1016/S0960-9822(02)00947-8

Illingworth, C. J. R. (2015). Fitness Inference from Short-Read Data: Within-Host Evolution of a Reassortant H5N1 Influenza Virus. *Molecular Biology and Evolution*, msv171. doi:10.1093/molbev/msv171

Illingworth, C. J. R. (2015). Fitness Inference from Short-Read Data: Within-Host Evolution of a Reassortant H5N1 Influenza Virus. *Molecular Biology and Evolution*, msv171. doi:10.1093/molbev/msv171

Ivankov, D. N., Finkelstein, A. V., & Kondrashov, F. a. (2014). A structural perspective of compensatory evolution. *Current Opinion in Structural Biology*, *26*(1), 104–112. doi:10.1016/j.sbi.2014.05.004

Jacquier, H., Birgy, A., Le Nagard, H., Mechulam, Y., Schmitt, E., Glodt, J., … Tenaillon, O. (2013a). Capturing the mutational landscape of the beta-lactamase TEM-1. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(32), 13067–72. doi:10.1073/pnas.1215206110

Jacquier, H., Birgy, A., Le Nagard, H., Mechulam, Y., Schmitt, E., Glodt, J., … Tenaillon, O. (2013b). Capturing the mutational landscape of the beta-lactamase TEM-1. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(32), 13067–13072. doi:10.1073/pnas.1215206110

Jelier, R., Semple, J. I., Garcia-Verdugo, R., & Lehner, B. (2011a). Predicting phenotypic variation in yeast from individual genome sequences. *Nature Genetics*, *43*(12), 1270–1274. doi:10.1038/ng.1007

Jelier, R., Semple, J. I., Garcia-Verdugo, R., & Lehner, B. (2011b). Predicting phenotypic variation in yeast from individual genome sequences. *Nature Genetics*, *43*(12), 1270–1274. doi:10.1038/ng.1007

Jones, C. D., Cox, P. M., Carmichael, G., Charlson, R. J., Rodhe, H., Crutzen, P. J., … Agency, I. D. (2009a). References and Notes 1., *323*(January), 2007–2010.

Jones, C. D., Cox, P. M., Carmichael, G., Charlson, R. J., Rodhe, H., Crutzen, P. J., … Agency, I. D. (2009b). References and Notes 1., *323*(January), 2007–2010.

Kimura, M., & Crow, J. F. (1978). Effect of overall phenotypic selection on genetic change at individual loci. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(12), 6168–6171. doi:10.1073/pnas.75.12.6168

Kimura, M., & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics*, *54*(6), 1337–1351. doi:10.1002/jlac.19666940111

Kondrashov, D. a, & Kondrashov, F. a. (2014a). Topological features of rugged fitness landscapes in sequence space. *Trends in Genetics*, *31*(1), 24–33. doi:10.1016/j.tig.2014.09.009

Kondrashov, D. a, & Kondrashov, F. a. (2014b). Topological features of rugged fitness landscapes in sequence space. *Trends in Genetics*, *31*(1), 24–33. doi:10.1016/j.tig.2014.09.009

Kouyos, R. D., Leventhal, G. E., Hinkley, T., Haddad, M., Whitcomb, J. M., Petropoulos, C. J., & Bonhoeffer, S. (2012a). Exploring the complexity of the HIV-1 fitness landscape. *PLoS Genetics*, *8*(3). doi:10.1371/journal.pgen.1002551

Kouyos, R. D., Leventhal, G. E., Hinkley, T., Haddad, M., Whitcomb, J. M., Petropoulos, C. J., & Bonhoeffer, S. (2012b). Exploring the complexity of the HIV-1 fitness landscape. *PLoS Genetics*, *8*(3). doi:10.1371/journal.pgen.1002551

Krishnakumar, R., & Ling, J. (2014). Experimental challenges of sense codon reassignment: An innovative approach to genetic code expansion. *FEBS Letters*, *588*(3), 383–388. doi:10.1016/j.febslet.2013.11.039

Krzycki, J. a. (2005). The direct genetic encoding of pyrrolysine. *Current Opinion in Microbiology*, *8*(6), 706–712. doi:10.1016/j.mib.2005.10.009

Kun, Á., & Szathmáry, E. (2015a). Fitness Landscapes of Functional RNAs. *Life*, *5*(3), 1497–1517. doi:10.3390/life5031497

Kun, Á., & Szathmáry, E. (2015b). Fitness Landscapes of Functional RNAs. *Life*, *5*(3), 1497–1517. doi:10.3390/life5031497

Kvitek, D. J., & Sherlock, G. (2011a). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genetics*, *7*(4). doi:10.1371/journal.pgen.1002056

Kvitek, D. J., & Sherlock, G. (2011b). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genetics*, *7*(4). doi:10.1371/journal.pgen.1002056

Lalić, J., & Elena, S. F. (2015a). The impact of high-order epistasis in the within-host fitness of a positive-sense plant RNA virus. *Journal of Evolutionary Biology*, n/a–n/a. doi:10.1111/jeb.12748

Lalić, J., & Elena, S. F. (2015b). The impact of high-order epistasis in the within-host fitness of a positive-sense plant RNA virus. *Journal of Evolutionary Biology*, n/a–n/a. doi:10.1111/jeb.12748

Laughlin, D. C., & Messier, J. (2015a). Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends in Ecology & Evolution*, 1–10. doi:10.1016/j.tree.2015.06.003

Laughlin, D. C., & Messier, J. (2015b). Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends in Ecology & Evolution*, 1–10. doi:10.1016/j.tree.2015.06.003

Lauring, A. S., & Andino, R. (2011a). Exploring the fitness landscape of an RNA virus by using a universal barcode microarray. *Journal of Virology*, *85*(8), 3780–3791. doi:10.1128/JVI.02217-10

Lauring, A. S., & Andino, R. (2011b). Exploring the fitness landscape of an RNA virus by using a universal barcode microarray. *Journal of Virology*, *85*(8), 3780–3791. doi:10.1128/JVI.02217-10

Lehner, B. (2011a). Molecular mechanisms of epistasis within and between genes. *Trends in Genetics*, *27*(8), 323–331. doi:10.1016/j.tig.2011.05.007

Lehner, B. (2011b). Molecular mechanisms of epistasis within and between genes. *Trends in Genetics*, *27*(8), 323–331. doi:10.1016/j.tig.2011.05.007

Lehner, B. (2013a). Genotype to phenotype: lessons from model organisms for human genetics. *Nature Reviews. Genetics*, *14*(3), 168–78. doi:10.1038/nrg3404

Lehner, B. (2013b). Genotype to phenotype: lessons from model organisms for human genetics. *Nature Reviews. Genetics*, *14*(3), 168–178. doi:10.1038/nrg3404

Lobanov, A. V, Turanov, A. a, Hatfield, D. L., & Gladyshev, V. N. (2010a). Dual functions of codons in the genetic code. *Critical Reviews in Biochemistry and Molecular Biology*, *45*(4), 257–265. doi:10.3109/10409231003786094

Lobanov, A. V, Turanov, A. a, Hatfield, D. L., & Gladyshev, V. N. (2010b). Dual functions of codons in the genetic code. *Critical Reviews in Biochemistry and Molecular Biology*, *45*(4), 257–265. doi:10.3109/10409231003786094

Lobkovsky, A. E., Wolf, Y. I., & Koonin, E. V. (2011a). Predictability of evolutionary trajectories in fitness landscapes. *PLoS Computational Biology*, *7*(12). doi:10.1371/journal.pcbi.1002302

Lobkovsky, A. E., Wolf, Y. I., & Koonin, E. V. (2011b). Predictability of evolutionary trajectories in fitness landscapes. *PLoS Computational Biology*, *7*(12). doi:10.1371/journal.pcbi.1002302

Lozovsky, E. R., Chookajorn, T., Brown, K. M., Imwong, M., Shaw, P. J., Kamchonwongpaisan, S., … Hartl, D. L. (2009). Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(29), 12025–12030. doi:10.1073/pnas.0905922106

Mackay, T. F. C. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nature Reviews. Genetics*, *15*(1), 22–33. doi:10.1038/nrg3627

Mackinnon, M. J., & Marsh, K. (2010a). The selection landscape of malaria parasites. *Science (New York, N.Y.)*, *328*(5980), 866–871. doi:10.1126/science.1185410

Mackinnon, M. J., & Marsh, K. (2010b). The selection landscape of malaria parasites. *Science (New York, N.Y.)*, *328*(5980), 866–871. doi:10.1126/science.1185410

Mahapatra, A., Patel, A., Soares, J. a., Larue, R. C., Zhang, J. K., Metcalf, W. W., & Krzycki, J. a. (2006). Characterization of a Methanosarcina acetivorans mutant unable to translate UAG as pyrrolysine. *Molecular Microbiology*, *59*(1), 56–66. doi:10.1111/j.1365-2958.2005.04927.x

Mann, J. K., Barton, J. P., Ferguson, A. L., Omarjee, S., Walker, B. D., Chakraborty, A., & Ndung'u, T. (2014a). The Fitness Landscape of HIV-1 Gag: Advanced Modeling Approaches and Validation of Model Predictions by In Vitro Testing. *PLoS Computational Biology*, *10*(8), e1003776. doi:10.1371/journal.pcbi.1003776

Mann, J. K., Barton, J. P., Ferguson, A. L., Omarjee, S., Walker, B. D., Chakraborty, A., & Ndung'u, T. (2014b). The Fitness Landscape of HIV-1 Gag: Advanced

Modeling Approaches and Validation of Model Predictions by In Vitro Testing. *PLoS Computational Biology*, *10*(8), e1003776. doi:10.1371/journal.pcbi.1003776

Manuscript, A. (2012a). NIH Public Access. *Changes*, *29*(6), 997–1003. doi:10.1016/j.biotechadv.2011.08.021.Secreted

Manuscript, A. (2012b). NIH Public Access. *Changes*, *29*(6), 997–1003. doi:10.1016/j.biotechadv.2011.08.021.Secreted

Manuscript, A. (2012c). NIH Public Access. *Changes*, *29*(6), 997–1003. doi:10.1016/j.biotechadv.2011.08.021.Secreted

Manuscript, A. (2012d). NIH Public Access. *Changes*, *29*(6), 997–1003. doi:10.1016/j.biotechadv.2011.08.021.Secreted

Manuscript, A. (2012e). NIH Public Access. *Changes*, *29*(6), 997–1003. doi:10.1016/j.biotechadv.2011.08.021.Secreted

Martin, G., Elena, S. F., & Lenormand, T. (2007a). Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature Genetics*, *39*(4), 555–560. doi:10.1038/ng1998

Martin, G., Elena, S. F., & Lenormand, T. (2007b). Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature Genetics*, *39*(4), 555–560. doi:10.1038/ng1998

Mccandlish, D. M. (2011a). Visualizing fitness landscapes. *Evolution*, *65*(6), 1544–1558. doi:10.1111/j.1558-5646.2011.01236.x

Mccandlish, D. M. (2011b). Visualizing fitness landscapes. *Evolution*, *65*(6), 1544–1558. doi:10.1111/j.1558-5646.2011.01236.x

McCandlish, D. M., Rajon, E., Shah, P., Ding, Y., & Plotkin, J. B. (2013). The role of epistasis in protein evolution. *Nature*, *497*(7451), E1–E2. doi:10.1038/nature12219

Meer, M. V, Kondrashov, A. S., Artzy-Randrup, Y., & Kondrashov, F. a. (2010a). Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature*, *464*(7286), 279–282. doi:10.1038/nature08691

Meer, M. V, Kondrashov, A. S., Artzy-Randrup, Y., & Kondrashov, F. a. (2010b). Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. *Nature*, *464*(7286), 279–282. doi:10.1038/nature08691

Meini, M.-R., Tomatis, P. E., Weinreich, D. M., & Vila, a. J. (2015). Quantitative Description of a Protein Fitness Landscape based on Molecular Features. *Molecular Biology and Evolution*. doi:10.1093/molbev/msv059

Melnikov, A., Rogov, P., Wang, L., Gnirke, A., & Mikkelsen, T. S. (2014a). Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Research*, *42*(14), 1–8. doi:10.1093/nar/gku511

Melnikov, A., Rogov, P., Wang, L., Gnirke, A., & Mikkelsen, T. S. (2014b). Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Research*, *42*(14), 1–8. doi:10.1093/nar/gku511

Milkman, R. (1978). Selection differentials and selection coefficients. *Genetics*, *88*(2), 391–403.

Nahum, J. R., Godfrey-Smith, P., Harding, B. N., Marcus, J. H., Carlson-Stevermer, J., & Kerr, B. (2015a). A tortoise–hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. *Proceedings of the National Academy of Sciences*, *112*(24), 201410631. doi:10.1073/pnas.1410631112

Nahum, J. R., Godfrey-Smith, P., Harding, B. N., Marcus, J. H., Carlson-Stevermer, J., & Kerr, B. (2015b). A tortoise–hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. *Proceedings of the National Academy of Sciences*, *112*(24), 201410631. doi:10.1073/pnas.1410631112

Namy, O., Zhou, Y., Gundllapalli, S., Polycarpo, C. R., Denise, A., Rousset, J. P., … Ambrogelly, A. (2007). Adding pyrrolysine to the Escherichia coli genetic code. *FEBS Letters*, *581*(27), 5282–5288. doi:10.1016/j.febslet.2007.10.022

Naumenko, S. a., Kondrashov, a. S., & Bazykin, G. a. (2012). Fitness conferred by replaced amino acids declines with time. *Biology Letters*, *8*(5), 825–828. doi:10.1098/rsbl.2012.0356

Neidhart, J., Szendro, I. G., & Krug, J. (2013a). Exact results for amplitude spectra of fitness landscapes. *Journal of Theoretical Biology*, *332*, 218–227. doi:10.1016/j.jtbi.2013.05.002

Neidhart, J., Szendro, I. G., & Krug, J. (2013b). Exact results for amplitude spectra of fitness landscapes. *Journal of Theoretical Biology*, *332*, 218–227. doi:10.1016/j.jtbi.2013.05.002

Neidhart, J., Szendro, I. G., & Krug, J. (2014a). Adaptation in Tunably Rugged Fitness Landscapes: The Rough Mount Fuji Model. *Genetics*, *198*(October), 1–43. doi:10.1534/genetics.114.167668

Neidhart, J., Szendro, I. G., & Krug, J. (2014b). Adaptation in Tunably Rugged Fitness Landscapes: The Rough Mount Fuji Model. *Genetics*, *198*(October), 1–43. doi:10.1534/genetics.114.167668

Olson, C. A., Wu, N. C., & Sun, R. (2014). A Comprehensive Biophysical Description of Pairwise Epistasis throughout an Entire Protein Domain. *Current Biology*, *24*(22), 2643–2651. doi:10.1016/j.cub.2014.09.072

Ostrowski, E. a, Ofria, C., & Lenski, R. E. (2015a). Genetically integrated traits and rugged adaptive landscapes in digital organisms. *BMC Evolutionary Biology*, *15*(1), 1–14. doi:10.1186/s12862-015-0361-x

Ostrowski, E. a, Ofria, C., & Lenski, R. E. (2015b). Genetically integrated traits and rugged adaptive landscapes in digital organisms. *BMC Evolutionary Biology*, *15*(1), 1–14. doi:10.1186/s12862-015-0361-x

Otwinowski, J., & Nemenman, I. (2013a). Genotype to Phenotype Mapping and the Fitness Landscape of the E. coli lac Promoter. *PLoS ONE*, *8*(5). doi:10.1371/journal.pone.0061570

Otwinowski, J., & Nemenman, I. (2013b). Genotype to Phenotype Mapping and the Fitness Landscape of the E. coli lac Promoter. *PLoS ONE*, *8*(5). doi:10.1371/journal.pone.0061570

Palmer, A. C., Toprak, E., Baym, M., Kim, S., Veres, A., Bershtein, S., & Kishony, R. (2015). Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nature Communications*, *6*(May 2014), 7385. doi:10.1038/ncomms8385

Parera, M., & Martinez, M. A. (2014). Strong epistatic interactions within a single protein. *Molecular Biology and Evolution*, *31*(6), 1546–1553. doi:10.1093/molbev/msu113

Paul, D. B. (1987). *"Our load of mutations" revisited. Journal of the History of Biology* (Vol. 20). doi:10.1007/BF00139458

Pitt, J. N., & Ferré-D'Amaré, A. R. (2010a). Rapid construction of empirical RNA fitness landscapes. *Science (New York, N.Y.)*, *330*(6002), 376–379. doi:10.1126/science.1192001

Pitt, J. N., & Ferré-D'Amaré, A. R. (2010b). Rapid construction of empirical RNA fitness landscapes. *Science (New York, N.Y.)*, *330*(6002), 376–379. doi:10.1126/science.1192001

Podgornaia, A. I., & Laub, M. T. (n.d.-a). Pervasive degeneracy and epistasis in a protein-protein interface.

Podgornaia, A. I., & Laub, M. T. (n.d.-b). Pervasive degeneracy and epistasis in a protein-protein interface.

Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M., & Tans, S. J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, *445*(7126), 383–386. doi:10.1038/nature05451

Poon, A., & Chao, L. (2005a). The rate of compensatory mutation in the DNA bacteriophage ??X174. *Genetics*, *170*(3), 989–999. doi:10.1534/genetics.104.039438

Poon, A., & Chao, L. (2005b). The rate of compensatory mutation in the DNA bacteriophage ??X174. *Genetics*, *170*(3), 989–999. doi:10.1534/genetics.104.039438

Povolotskaya, I. S., & Kondrashov, F. a. (2010a). Sequence space and the ongoing expansion of the protein universe. *Nature*, *465*(7300), 922–926. doi:10.1038/nature09105

Povolotskaya, I. S., & Kondrashov, F. a. (2010b). Sequence space and the ongoing expansion of the protein universe. *Nature*, *465*(7300), 922–926. doi:10.1038/nature09105

Povolotskaya, I. S., Kondrashov, F. a, Lleda, A., & Vlasov, P. K. (2012). Stop codons in bacteria are not selectively equivalent. *Biology Direct*, *7*(1), 30. doi:10.1186/1745-6150-7-30

Reynolds, K. a., McLaughlin, R. N., & Ranganathan, R. (2011). Hot spots for allosteric regulation on protein surfaces. *Cell*, *147*(7), 1564–1575. doi:10.1016/j.cell.2011.10.049

Rogozin, I. B., Thomson, K., Csürös, M., Carmel, L., & Koonin, E. V. (2008). Homoplasy in genome-wide analysis of rare amino acid replacements: the molecular-evolutionary basis for Vavilov's law of homologous series. *Biology Direct*, *3*, 7. doi:10.1186/1745-6150-3-7

Roscoe, B. P., Thayer, K. M., Zeldovich, K. B., Fushman, D., & Bolon, D. N. a. (2013). Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *Journal of Molecular Biology*, *425*(8), 1363–1377. doi:10.1016/j.jmb.2013.01.032

Sabchareon, a., Burnouf, T., Ouattara, D., Attanath, P., Bouharoun-Tayoun, H., Chantavanich, P., … Druilhe, P. (1991a). Parasitologic and clinical human response to immunoglobulin administration in falciparum malaria. *American Journal of Tropical Medicine and Hygiene*, *45*(3), 297–308. doi:10.1155/2010/453642

Sabchareon, a., Burnouf, T., Ouattara, D., Attanath, P., Bouharoun-Tayoun, H., Chantavanich, P., … Druilhe, P. (1991b). Parasitologic and clinical human response to immunoglobulin administration in falciparum malaria. *American Journal of Tropical Medicine and Hygiene*, *45*(3), 297–308. doi:10.1155/2010/453642

Sandler, I., Zigdon, N., Levy, E., & Aharoni, A. (2014). The functional importance of co-evolving residues in proteins. *Cellular and Molecular Life Sciences*, *71*(4), 673–682. doi:10.1007/s00018-013-1458-2

Schimmel, P., & Beebe, K. (2014). News and views. *International Wound Journal*, *11*(4), 344–7. doi:10.1111/iwj.12332

Seetharaman, S., & Jain, K. (2014a). Length of adaptive walk on uncorrelated and correlated fitness landscapes, *032703*(September), 1–12. doi:10.1103/PhysRevE.90.032703

Seetharaman, S., & Jain, K. (2014b). Length of adaptive walk on uncorrelated and correlated fitness landscapes, *032703*(September), 1–12. doi:10.1103/PhysRevE.90.032703

Shnol, E. E., & Kondrashov, a. S. (1993). The effect of selection on the phenotypic variance. *Genetics*, *134*(3), 995–996.

Smith, J. M. (1970). Natural selection and the concept of a protein space. *Nature*, *225*(5232), 563–564. doi:10.1038/225563a0

Soskine, M., & Tawfik, D. S. (2010). Mutational effects and the evolution of new protein functions. *Nature Reviews. Genetics*, *11*(8), 572–582. doi:10.1038/nrg2808

Srinivasan, G., James, C. M., & Krzycki, J. a. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science (New York, N.Y.)*, *296*(5572), 1459–1462. doi:10.1126/science.1069588

Stiffler, M. A., Hekstra, D. R., & Ranganathan, R. (2015). Evolvability as a Function of Purifying Selection in TEM-1 β-Lactamase. *Cell*, *160*(5), 882–892. doi:10.1016/j.cell.2015.01.035

Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., Kondrashov, a S., & Bork, P. (2001). Prediction of deleterious human alleles. *Human Molecular Genetics*, *10*(6), 591–597. doi:10.1093/hmg/10.6.591

Szamecz, B., Boross, G., Kalapis, D., Kovács, K., Fekete, G., Farkas, Z., … Pál, C. (2014a). The Genomic Landscape of Compensatory Evolution. *PLoS Biology*, *12*(8), e1001935. doi:10.1371/journal.pbio.1001935

Szamecz, B., Boross, G., Kalapis, D., Kovács, K., Fekete, G., Farkas, Z., … Pál, C. (2014b). The Genomic Landscape of Compensatory Evolution. *PLoS Biology*, *12*(8), e1001935. doi:10.1371/journal.pbio.1001935

Taylor, M. B., & Ehrenreich, I. M. (2015). Higher-order genetic interactions and their contribution to complex traits. *Trends in Genetics*, *31*(1), 34–40. doi:10.1016/j.tig.2014.09.001

Theil Have, C., Zambach, S., & Christiansen, H. (2013). Effects of using coding potential, sequence conservation and mRNA structure conservation for predicting pyrroly-sine containing genes. *BMC Bioinformatics*, *14*(1), 118. doi:10.1186/1471-2105-14-118

Théobald-Dietrich, A., Giegé, R., & Rudinger-Thirion, J. (2005). Evidence for the existence in mRNAs of a hairpin element responsible for ribosome dependent pyrrolysine insertion into proteins. *Biochimie*, *87*(9-10), 813–817. doi:10.1016/j.biochi.2005.03.006

Wan, W., Tharp, J. M., & Liu, W. R. (2014). Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1844*(6), 1059–1070. doi:10.1016/j.bbapap.2014.03.002Review

Wei, W.-H., Hemani, G., & Haley, C. S. (2014). Detecting epistasis in human complex traits. *Nature Reviews Genetics*, *15*(11), 722–733. doi:10.1038/nrg3747

Weinreich, D. M., & Knies, J. L. (2013). Fisher's Geometric Model Of Adaptation Meets The Functional Synthesis: Data On Pairwise Epistasis For Fitness Yields Insights Into The Shape And Size Of Phenotype Space. *Evolution*, *67*(10), 2957–2972. doi:10.1111/evo.12156

Weinreich, D. M., Lan, Y., Wylie, C. S., & Heckendorn, R. B. (2013). Should evolutionary geneticists worry about higher-order epistasis? *Current Opinion in Genetics and Development*, *23*(6), 700–707. doi:10.1016/j.gde.2013.10.007

Weinreich, D. M., Watson, R. a, & Chao, L. (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution; International Journal of Organic Evolution*, *59*(6), 1165–1174. doi:10.1111/j.0014-3820.2005.tb01768.x

Woo, H. J., & Reifman, J. (2014a). Quantitative Modeling of Virus Evolutionary Dynamics and Adaptation in Serial Passages Using Empirically Inferred Fitness Landscapes. *Journal of Virology*, *88*(2), 1039–1050. doi:10.1128/JVI.02958-13

Woo, H. J., & Reifman, J. (2014b). Quantitative Modeling of Virus Evolutionary Dynamics and Adaptation in Serial Passages Using Empirically Inferred Fitness Landscapes. *Journal of Virology*, *88*(2), 1039–1050. doi:10.1128/JVI.02958-13

Wright, S. (1932a). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress on Genetics*. doi:citeulike-article-id:1586133

Wright, S. (1932b). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress on Genetics*. doi:citeulike-article-id:1586133

Zhang, W., Dourado, D. F. a. R., Fernandes, P. A., Ramos, M. J., & Mannervik, B. (2012a). Multidimensional epistasis and fitness landscapes in enzyme evolution. *Biochemical Journal*, *445*(1), 39–46. doi:10.1042/BJ20120136

Zhang, W., Dourado, D. F. a. R., Fernandes, P. A., Ramos, M. J., & Mannervik, B. (2012b). Multidimensional epistasis and fitness landscapes in enzyme evolution. *Biochemical Journal*, *445*(1), 39–46. doi:10.1042/BJ20120136

Zhang, Y., Baranov, P. V., Atkins, J. F., & Gladyshev, V. N. (2005a). Pyrrolysine and selenocysteine use dissimilar decoding strategies. *Journal of Biological Chemistry*, *280*(21), 20740–20751. doi:10.1074/jbc.M501458200

Zhang, Y., Baranov, P. V., Atkins, J. F., & Gladyshev, V. N. (2005b). Pyrrolysine and selenocysteine use dissimilar decoding strategies. *Journal of Biological Chemistry*, *280*(21), 20740–20751. doi:10.1074/jbc.M501458200

Zhang, Y., & Gladyshev, V. N. (2007a). High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis. *Nucleic Acids Research*, *35*(15), 4952–4963. doi:10.1093/nar/gkm514

Zhang, Y., & Gladyshev, V. N. (2007b). High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis. *Nucleic Acids Research*. doi:10.1093/nar/gkm514

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012a). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, *109*(4), 1193–1198. doi:10.1073/pnas.1119675109

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012b). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, *109*(4), 1193–1198. doi:10.1073/pnas.1119675109

9. Appendix

1: Koblik EA, Red'kin YA, **Meer MS**, Derelle R, Golenkina SA, Kondrashov FA, Arkhipov VY. Acrocephalus orinus: a case of mistaken identity. PLoS One. 2011 Apr 22;6(4):e17716. doi: 10.1371/journal.pone.0017716. PubMed PMID: 21526114; PubMed Central PMCID: PMC3081296.

2: Arkhipova OV, **Meer MV**, Mikoulinskaia GV, Zakharova MV, Galushko AS, Akimenko VK, Kondrashov FA. Recent Origin of the Methacrylate Redox System in Geobacter sulfurreducens AM-1 through Horizontal Gene Transfer. PLoS One. 2015 May 11;10(5):e0125888. doi: 10.1371/journal.pone.0125888. eCollection 2015. PubMed PMID: 25962149; PubMed Central PMCID: PMC4427408.

Koblik EA, Red'kin YA, Meer MS, Derelle R, Golenkina SA, Kondrashov FA, Arkhipov VY. Acrocephalus orinus: a case of mistaken identity. PLoS One. 2011 Apr 22;6(4):e17716. doi: 10.1371/journal.pone.0017716.

Arkhipova OV, Meer MV, Mikoulinskaia GV, Zakharova MV, Galushko AS, Akimenko VK, Kondrashov FA. Recent Origin of the Methacrylate Redox System in Geobacter sulfurreducens AM-1 through Horizontal Gene Transfer. PLoS One. 2015 May 11;10(5):e0125888. doi: 10.1371/journal.pone.0125888.