Departament de Ciències Experimentals i de la Salut
Universitat Pompeu Fabra

GENETIC VARIATION IN HUMANS AND CHIMPANZEES IN THE PRION PROTEIN GENE

TESI DOCTORAL

Marta Soldevila i Trepat

Sabadell, Abril del 2005

Departament de Ciències Experimentals i de la Salut
Universitat Pompeu Fabra

GENETIC VARIATION IN HUMANS AND CHIMPANZEES IN THE PRION PROTEIN GENE

Marta Soldevila i Trepat

Memòria presentada per optar al títol de Doctora en Biologia per la Universitat Pompeu Fabra. Aquest treball ha estat realitzat sota la direcció del Dr. Jaume Bertranpetit i Busquets a la Unitat de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra.

Jaume Bertranpetit i Busquets
Director

Marta Soldevila i Trepat

Sabadell, Abril del 2005

Al Carles.

A tota la meva família,
i molt especialment
a ma germana, l´Anna.

Als tots els companys i amics
que m´han ajudat a fer-ho possible.

Els prions representen, en realitat, el triomf de la investigació científica sobre el prejudici.

S.B. Prusiner -Premi Nobel

CONTENTS

AGRAÏMENTS

Haig d´admetre que aquest, els agraïments, ha estat el primer apartat de la tesi que he escrit. No només és el que em feia més il·lusió, sinó que juntament amb el treball realitzat el component humà ha estat un factor clau per disfrutar d´aquests anys. Gràcies a tots pel vostre suport.

I, com que sé que la secció és la més llegida, parlant amb termes científics, l´impact factor de la secció agraïments és sempre molt més elevat que la resta de la tesi...., doncs espero que la disfruteu!

M´agradaria agrair molt especialment al meu director i mestre, Jaume Bertranpetit, per aquests temps de dedicació, de supervisió, de ciència, d´amistat i de política. Ressaltar la meva admiració per la lluita per estar al dia de tot el que es mou en el camp científic (Jaume, no sé com aguantes tantes reunions!), i el seu notable interès per la cultura en general, informant-nos constantment de les exposicions, concerts, llibres... més interessants de cada setmana. Gràcies per deixar-me compartir un tros de la teva vida!.

També m´agradaria donar les gràcies al Francesc Calafell per la seva inestimable ajuda en els anàlisis que he realitzat, i també per ser aquella persona que fent una pregunta, de qualsevol tema, sabia donar-me una resposta o com a mínim sabia indicar-me a on trobar-la. Del David sempre recordaré la seva alegria i les seves bromes.... i el trio!. Ostres, al final et trobaré fins i tot a faltar...!!. Recorda: no facis res que jo no faria!. De l´Elena

la seva disciplina, la seva constància... el seu ordre i per ser la meva companya de viatge (gràcies per ser puntual als llocs, i no fer-me sentir diferent!). I pels molts bons moments que hem compartit!. Que segur que seràs una mare excel·lent!. De l´Arcadi, recordaré moltes coses, les de la UAB i les de la UPF, però per sobre de tot, recordaré que ets sempre allà quan et necessito. Petons a l´Adela!. De l´Aida, com no, recordaré els moments compartits, i especialment els ànims que em transmetia. Estic segura que faràs una bona carrera científica, endavant!. De l´Anna G, la G, per l´energia i les nostres converses catalano-madrilenyes... i sobretot per escoltar-me quan calia. Del Jordi, buff, la seva música, els seus cants d´òpera, i els seus mails dels usa. Ànim nen, que ets el rei de la neurologia!. De la Mònica, trobaré a faltar el seu costat sentimental que descobreixes poc a poc. Quina llàstima no haver-lo descobert abans.... cuida´t molt eh!. Apa, que ja tens pis!!! Per fi!. De la R, de ramírez, pels temps complicats que hem compartit juntes, per les infinites converses, múltiples anàlisis (mil gràcies! + mil gràcies més) i vaja, que hagis fet que hagi estat una mica menys sola fent ciència i una mica més feliç. Que ja és molt. Cuida´t i cuida el Roger!. De la Gemma, GEM!, trobaré a faltar moltes coses (SO MANY!!), però potser el fet que sigui una persona que m´hagi alegrat molts dies amb les seves històries increïbles. Ah i també agrair-te les teves correccions del meu anglès en aquesta tesi. Petons a l´Álex en gallego-portuguès del teu. De la Cristina, uff, Cris, tu i jo com a mínim ja hauríem de portar escrits dos llibres! Jo crec que dues tesis sortirien, no?. Vigila la serotonina. Sigues feliç. De la Lourdes trobaré a faltar la seva companyia, i sobretot, la seva paciència infinita, i la seva bona voluntat. Prometo venir a

Panticosa!!!!. De la Michi, la seva capacitat per fer-me riure, de les seves bromes intel·ligents, la seva il·lusió per les coses....I moltíssimes gràcies per les teves correccions del meu anglès en aquesta tesi!!!! Many thanks. Or in irish pronunciation, many tanks. All the best and don´t worry you will get what you want!. De l´Òscar, oscarín, ets un tio fantàstic, bona persona de veritat. Gràcies per l´ajuda quan te l´he demanada. Fas la ciència fàcil al teu costat. Al Tomàs, pels moments bons que m´has fet passar, per la nostra connexió ràpida, per ser com ets, i per tenir una xicota tan maca, la Núria. I pel molt que m´has ajudat amb els teus programes. Gràcies. A l´Stephi, per la seva veu relaxant i sempre educada, gràcies per recolzar-me. Pel seu somriure constant després de la teva tesi...Et desitjo el millor. A l´Andrés, per ser un bon alumne de català, com ella, i deixar-me fer bromes sobre els mexicans...sense enfadar-te!. Hasta puede que te extrañe ;-P. A l´Anna petita, per la seva rialla i per recordar-me la meva infantesa amb les seves històries de l´esplai. Al Ferran, per haver-me fet sempre bona cara, fins i tot quan era una estudiant que anava per allà el departament de la UAB, i per ser sempre tan simpàtic i "acceptar" les meves correccions dels seus paranys lingüístics...no canviïs!. Que si et trobaré a faltar? Yesssssssssss (i moltes més esses). Una abraçada a la Mireia. A la Pérez, que m´ha fet riure molt amb les seves històries...així sempre una mica pujades de to o una mica increïbles, però amb molta gràcia. Records a l´Eduard. Al Roger, simpàtic, guapet, vaja, del 77... i per la seva capacitat de ser feliç i entendre de qualsevol cosa. I als nous, o no tant nous: Morcillo, pel teu somriure constant i les teves bromes ...al Pep, per explicar tot el que amaguen les màquines de forma senzilla, a l´Elodie, per recordar-me que algun dia vull aprendre francès i

per aquells dies que sembla que s´hagi pres alguna cosa al matí que fa que rigui tota l´estona ...jo crec que té alguna pastilla francesa que li produeix efectes secundaris..Ah! i prometo fer classes de català més pràctiques que en aquest arbre hi ha un esquirol! Espero no fer....deux fois trop tard !., a la Kristin, i al seu home sabadellenc, el Salvador, que ens heu acompanyat encantats moltes vegades a Sabadell i ens heu fet recordar els vells temps a Islàndia. Takk. I a l´ última incorporació, la Karla, que sempre m´ha dedicat un somriure i que un dia em va regalar un punt de llibre que encara guardo...Gràcies. Quien va a recordarte tan bien las horas de comer???. I al Marc... perquè sempre m´ha parlat amb sinceritat i perquè estic segura que les coses li aniran millor d´aquí a un temps...Molta sort!. I als que vindran al grup mentre jo dipositi la tesi i la llegeixi...

Al marmi, o marmí, per ser un tros de pa de Berga, i per compartir bons moments a l´autònoma amb la nostra amiga Meri, al Flope que a part de ser veïns, vam compartir moltes històries (ei, no penseu malament), a l´Olga, amb qui de seguida vaig connectar, *rainbow* et trobaré a faltar, l´Ainhoa, per les seves aventures al camp, el Bruno, per pujar-me l´ego amb els seus comentaris, la Montse, tot i que m´hauria agradat estar més temps juntes, a tots...una abraçada ben gran. I al Xavi, com no, per parlar bé de mi al Carles, quan va ser necessari (ja m´entens!).

A la gent que ha passat per aquí al lab, el Jesús (llàstima que només hagin estat uns mesos), la Chiara, l´Arnau, la Valentina, la Christine, la Heidi, ...

Als que han compartit el nostre lab, els de bioquímica Miquel Àngel, Xavi i Pep, que ens ha ensenyat que és fer protocols llargs de veritat!. Sort Xavi en el teu nou futur…

Agrair també als membres de l´Hospital Clínic, Albert Saiz, Carlos Nos, Raquel Sánchez-Valle, la Fina i la Susanna i molt especialment a Jordi Yagüe, per tenir l´oportunitat de treballar amb les mostres de pacients, per facilitar-me en tot moment la feina i per contestar sempre de bon grat les meves preguntes i fer-m´ho una mica tot més fàcil.

A més voldria agrair a la gent del CreSA, entre ells el Martí Pumarola i la Montse Ordoñez, per ajudar-me amb la part no genètica de la meva tesi, i deixar-me entrar en el seu laboratori per a realitzar algunes proves amb el test PRIONICS. Gràcies.

A l´Antoine Blancher, que ens va ajudar molt a obtenir unes bones mostres de ximpanzé, que em van anar molt bé per al segon estudi, i pels seus comentaris.

Als que no saben que m´han ajudat, als pacients anònims, a qui em sap molt greu no haver pogut ajudar prou i als ximpanzés anònims, que potser arribarem a extingir-los abans que nosaltres puguem contestar la pregunta bàsica ...què ens diferencia d´ells... (espero que ens aguanteu molts més anys!)

Als nens respectius, Marçal, Mei-Xen, Adrià, Laia, Duna….que m´heu fet passar bones estones....sobretot tu Mei-Xen amb els fantàstics pentinats que m´has

fet..i a tu Laia, per deixar-te fer petons...I als que estan de camí...com el Rogeret.

Al Pablo, que ha compartit amb mi els últims mesos la difícil tasca de comprovar les seqüències que vaig generar a decode,...i moltes bones estones. Gràcies.

Als del grup del Guigó en general, i en particular al Sergi, per tantes coses, per escoltar-me, per entendre´m, per ajudar-me, per animar-me, per informar-me...Gràcies, de veritat. Gràcies per la tesi i gràcies també pel cap d´any. No l´he oblidat. Et trobaré a faltar. Estic segur que seràs un gran científic...si és que no ho ets encara. Al Pep i al Genís, que potser ja hauran publicat un altre science o nature, amb el temps d´escriure els agraïments, i per dedicar-me sempre uns quants somriures. Gràcies nois.

A tots aquells que vaig compartir alguns moments quan vaig estar a Islàndia, i en particular, la Catalina (i os estamos esperando en Barcelona!), la Malala (bueno, felicidades, ya sabes.., y mucha suerte), el Joan (a qui trobo a faltar una miqueta! i a qui li desitjo el millor), el Jesús (que me trató súper bien), a Emilio (mi coleguita), ...I els companys que a més em van ajudar, la Sigrun, les noies de l´Agnar.... i com no, a l´Agnar i a tota la seva família, especialment a la seva dona Anna i els nens, Guðrún Diljá, i Atli. Thank you a lot!. Takk fyrir síðast.

Als que han envoltat el grup, l´Òscar Fornas, la Damiana, la Raquel, l´Anna, la Laura....el Magano, també una forta abraçada.
I els que vaig conèixer a l´Autònoma, Hafid, Pepi, Bàrbara, Mario, Marta, Ester, Alfredo, Òscar, i tots els del grup de mutagènesi, especialment a la Sílvia (ara també coneguda com la tieta) i al Noel. I com no, els de la UB (i els ex de la UB), especialment amb els que he compartit un dinar, o un sopar, o una festa.. o alguna conversa, o algun moment de la meva tesi...

I molt especialment, a tota la meva família, als meus pares (perquè això és el fruit d´estudiar tants i tants anys...), germans i avis (que sempre m´han considerat més llesta del què realment sóc), i especialment ma mare i ma germana, que potser van entendre alguna cosa del què feia. Gràcies Anna pel teu recolzament, i per ser també una amiga.
I a la família del Carles, que havent-hi tants biòlegs, han entès bastant a què dedicava tants caps de setmana...i a tota la resta, per l´interès i reconeixement que m´han mostrat. Gràcies.

Als meus amics i amigues de sempre, especialment a la Sandra (que per a tu també han estat 4 anys i mig difícils i en els que has hagut de lluitar mooolt, ENDAVANT !), Olga (estic segura que aconseguiràs el que desitges), Gemma i Miquel (per les nostres converses de tren, i per compartir moltes altres coses.. i per les teves llàgrimes en un dia important per a mi), Meri (perquè tot i estar força lluny, has estat sempre a prop, ah! i una abraçada al Xevi), la Vanessa (prometo venir-te a veure més…quan hagi acabat amb això !), i la Cristina G. i

família (que ha aconseguit molt més que una tesi...) ....que m´heu posat les coses fàcils i m´heu animat i escoltat quan convenia. Moltes gràcies !!.

I finalment (i com dirien els anglesos, the last but not the least), dedicar la tesi al Carles, que m´ha canviat la vida, que m´ha ajudat quan convenia en les meves crisis personals i científiques, que m´ha donat estabilitat de tot tipus i a qui continuaré estimant durant molts i molts anys, molts més que el què ha durat aquesta tesi. Gràcies pels teus llibres, els teus quadres, les dades històriques, la cultura. Gràcies per la teva ètica. Gràcies per estimar-me. Gràcies pel dia 10, pel 22, pel 15 i pel viatge a Venècia, a Sri Lanka i Maldives, i pels viatges al sud de França....però sobretot, sobretot, pel dia a dia, que ha fet que hagi estat un plaer arribar a casa.... Gràcies per tot rei!.

Abans d´acabar però m´agradaria mencionar el fet que he tingut algunes "*batalles"* importants amb programes que no han acabat de funcionar, i amb tota aquella gent que han fet exactament el mateix que jo (vaja, "la competència") (tot i que cal dir-ho, són bona gent i els *reis* del camp: el Simon Mead i el John Collinge, a Londres i el John Hardy, el Martin Kreitman i l´Anna Di Rienzo, als USA)... que han aconseguit posar-me el nivell moooolt alt, encara que també m´han donat alguna alegria. També m´agradaria recordar, que he compartit amb ells bones idees, i un interès incessant en el tema de prions ...i el canibalisme. Ara, per una *becària* que comença... potser ha estat una mica massa...

Aquest treball l´he dut a terme mentre gaudia d´una beca de formació de personal investigador de la Generalitat de Catalunya (2001 FI 00632), de gener de 2001 fins a desembre de 2004. També gràcies a les ajudes de la Generalitat he pogut gaudir de tres borses de viatge que m´han permès realitzar part del meu treball a Islàndia. A més el treball ha comptat amb les subvencions de la Dirección General de Investigación Científica y Técnica (BMC 2001-0772) i de la Generalitat de Catalunya com a Grup de Recerca Consolidat (2001SGR00285). També voldria agrair a la Pompeu Fabra per afavorir la presentació de treballs d´aquesta tesi a congressos internacionals gràcies als seus programes PRESENTA i EBES que em foren adjudicats.

PERSPECTIVA GENERAL

TREBALL PRESENTAT AL PREMI COMUNICACIÓ CIENTÍFICA
"JOAN LLUÍS VIVES"
DESEMBRE 2003

DE LES VAQUES BOGES AL CANIBALISME....

Marta Soldevila
Unitat de Biologia Evolutiva, UPF

Quan el Dr. Carleton Gajdusek, guanyador d´un premi Nobel l´any 1976, se´n va anar a Papua Nova Guinea a mitjans de segle passat per estudiar una estranya malaltia que patien un grup denominat Fore, mai no hauria imaginat que algú, algun dia, escriuria sobre ell amb un títol com aquest. Jo per descomptat, tampoc. Però per difícil que sembli, tot té un sentit i una lògica.

Vaques boges i la revolució mediàtica

El fenòmen de les vaques boges començà ara farà vint anys, l´any 1984, quan un granger de Sussex (Anglaterra) va comunicar al govern anglès que una de les seves vaques es comportava de manera molt estranya, tremolava i no coordinava els seus moviments. Un any més tard es confirmava que aquest era el primer cas de BSE (del seu nom científic, *bovine spongiform encephalopaty*), una malaltia desconeguda fins llavors en vaques, anomenada popularment com "la malaltia de les vaques boges".

Però quina era la causa del sorgiment d´aquesta malaltia en les vaques? Es coneixia ja des del segle XVIII, com una malaltia semblant feia estralls en les ovelles. Afectava els ramats de França, on l´anomenaven la *tremblante*, o

tremolor, d´Alemanya, on rebé el nom de *Gnubberkrankheit*, o malaltia del picor; més tard va arribar als ramats d´Islàndia, on es denominà *Rida* (o descordinació del moviment, ataxia). Aquests noms indiquen, al mateix temps, les característiques clíniques principals associades a la malaltia. Però més endavant es conegué que aquesta malaltia en ovelles era endèmica en molts països, entre ells, el Regne Unit. I d´Escòcia ens ve el nom *Scrapie*, que és com se la defineix en la terminologia científica.

L´*scrapie*, va "saltar la barrera d´espècie" i passà de les ovelles a les vaques, i d´aquestes, saltà als humans. Aquest "salt" es produí per un canvi en el protocol en l´el·laboració de farines que es donaven a les vaques i que tenien un alt contingut proteic. Les farines contenien restes de vísceres d´ovelles, algunes d´elles infectades amb *scrapie.* Aquestes ja s´havien utilitzat durant les últimes dècades, però des dels anys vuitanta les condicions a què se sotmetien eren menys restrictives, i la proteïna causant de l´*scrapie* no s´inactivava amb aquest nou mètode.

No fou fins el 1988 que es prohibí en el Regne Unit alimentar a ruminants amb farines de contingut càrnic, i a més a més s´ordenà sacrificar per primera vegada totes les vaques infectades. A partir d´aquest any, es començaren a succeir els casos de la malaltia (BSE) a fora del Regne Unit, per posteriorment acabar trobant casos a tot Europa i fins i tot fora de les seves fronteres. Aquesta expansió de la malaltia es feu mitjançant l´exportació de boví i de "farines" britàniques infectades. El 1990, i quan començava a perillar la imatge del Regne Unit, el ministre d´Agricultura John Gummer, va intentar calmar les pors sobre la seguretat de la vedella menjant-se davant de la premsa de tot el

món, una hamburguesa junt amb la seva filla, Cordèlia, a les escales del Parlament.

Uns anys després, el 1995, morí Stephen Churchill, la primera víctima de la variant de la malaltia de Creutzfeldt-Jakob. Així fou com se l´anomenà en humans, però com que una malaltia molt similar i també d´origen priònic ja existia, es denominà *variant*, per tal de diferenciar-la de la resta. Així la variant es deu exclusivament a l´adquisició de la malaltia a través del consum de vedella amb BSE. Així, amb aquest cas, es demostrava que la proteïna priònica, causant de la malaltia, podia passar de la vedella a l´espècie humana. Per frenar aquesta malaltia s´ordenà eliminar milions de caps de vacú al Regne Unit. A Espanya la malaltia de les vaques boges no arribà fins l´any 2000, i ha anat en augment, ja que el 2003 ha resultat ser l´any en el qual s´han detectat més casos (uns 150), sumant en total més de 350 caps de vedella.

Creutzfeldt-Jakob (CJD): la malaltia en humans

El nom d´aquesta malaltia en humans fou introduit per W. Spielmeyer (1922), a partir dels casos publicats per dos metges alemanys de principis del segle passat, H. G. Creutzfeldt (1920) i A. Jakob (1921). Es tracta d´un rar trastorn del cervell, de tipus degeneratiu i invariablement mortal. Les malalties priòniques es classifiquen en tres categories: la forma hereditària, l´esporàdica, i l´adquirida.

La variant de CJD, produïda per consum de vedella amb BSE, es troba classificada dins del CJD adquirit, on comparteix la categoria amb altres formes, com el *kuru* (tremolor, en llengua indígena) de Nova Guinea. La forma

adquirida representa només un 5% dels casos totals, mentre que la forma hereditària es presenta en un 15% dels pacients i la resta, un 80%, s´agrupa en la categoria de formes esporàdiques. Les tres categories es diferencien en com es genera la malaltia. Així, la forma hereditària està causada per mutacions en el gen PRNP, el gen que codifica per la proteïna priònica, i pot rebre noms diferents segons la mutació i els símptomes dels pacients. Els noms que reben les formes genètiques són: Creutzfeldt-Jakob hereditari, Insomni Fatal Familiar (FFI), o la malaltia de Gerstmann-Straussler (GSS). En la forma adquirida, en canvi, els pacients no presenten mutacions en el gen, sinó que s´adquireix la malaltia a través del contacte amb una proteïna externa infecciosa. I finalment la forma esporàdica, en la qual s´agrupen tots aquells casos on tampoc no hi ha cap mutació en el gen PRNP ni hi ha hagut cap contacte amb la proteïna infecciosa, i que normalment es manifesta a partir d´una edat avançada. La seva causa segueix essent un misteri.

La importància de ser heterozigot: el kuru com a exemple.

El kuru és una malaltia d´origen priònic i de tipus adquirit. Aquesta malaltia va cridar l´atenció als anys cinquanta quan en una regió de Papua Nova Guinea es va detectar aquesta malaltia. Aquesta només afectava part d´una comunitat lingüística anomenada Fore, que estava aïllada geogràficament i reproductivament d´altres comunitats de la mateixa illa. Va ser la pràctica que van dur a terme aquestes comunitats, on familiars dels morts, mitjançant rituals tribals, consumien vísceres (entre elles cervell) dels seus relatius quan aquests morien. Aquesta pràctica però, era relativament recent, ja que va començar a

partir del segle XIX. El primer cas de kuru fou al voltant de l´any 1920 i la malaltia ràpidament va augmentar d´incidència. Les dones adultes i els nens d´ambdós sexes foren afectats primàriament, reflectint la seva exposició selectiva en aquest tipus de rituals –on en canvi, els homes adults participaven poc. En el seu punt màxim, aquesta malaltia va matar l´1% de la població Fore cada any, i alguns pobles es van quedar gairebé sense dones joves. Aquesta malaltia fou la primera de tipus priònic que mostrava clarament la seva transmissiblitat, i s´hipotetitza que el kuru s´originà pel consum inicial d´un individu amb CJD esporàdic. Així, a partir d´aquest, i a través dels seus familiars, s´haurien anat transmetent la malaltia, i com que són comunitats molt endogàmiques, aquesta s´havia estès molt fàcilment. Al prohibir aquesta pràctica, les autoritats australianes, en els anys cinquanta, la incidència del kuru es reduí espectacularment, fins arribar a la desaparició d´aquest en les noves generacions.

Va ser en un estudi recent que es veié que en molt poc temps les poblacions afectades per la malaltia havien estat sotmeses a una selecció brutal a favor dels heterozigots (a tenir dos variants diferents, metionina i valina en els dos cromosomes, un del pare i l´altre el provenint de la mare) pel codó 129 del gen PRNP. Això es va veure comparant els genotips d´aquestes poblacions sotmeses a aquest tipus de rituals, amb els genotips de la mateixa regió però que no havien estat exposades. I el resultat fou sorprenent, ja que les freqüències d´aquestes variants eren significativament diferents, i les poblacions que havien estat sotmeses tenien una freqüència molt més alta d´heterozigots que la població no exposada. Així es veia clarament que ser

heterozigot en la posició 129 del gen donava un avantatge selectiu extraordinari, i que d´alguna manera la selecció ens estava donant pistes de quin era el genotip més adient per a superar aquella situació. En el mateix estudi, es conclou que, estudiant poblacions de tot el món, només és possible d´explicar les característiques particulars d´aquest gen si hi ha hagut fenòmens similars als del kuru, i per tant, que el canibalisme segurament era una pràctica més freqüent del què ens pensàvem en el passat. Per altres estudis es coneix que aquesta selecció (anomenada selecció equilibradora) ha estat la més forta fins ara documentada en les poblacions humanes.

Test diagnòstic i variació humana

En el nostre grup vàrem veure l´oportunitat de desenvolupar un test diagnòstic per les malalties priòniques amb base hereditària en humans. Com hem comentat, diferents mutacions produeixen diferents malalties que tenen alguns símptomes diferenciats. Totes aquestes mutacions es troben únicament en un gen, un dels 25.000 que tenim en l´espècie humana. Aquest gen (PRNP) es troba en el cromosoma 20, i mitjançant els processos de transcripció i traducció es forma la proteïna priònica, que és una proteïna que de forma natural està implicada en la comunicació entre neurones del cervell (anomenada sinapsi neuronal), mitjançant la regulació del coure. És per això que aquesta proteïna es troba principalment (encara que no exclusivament) al cervell. Però, quines propietats té aquesta proteïna que la facin tan especial? Encara hi ha molts forats negres per resoldre pel què fa a aquestes propietats, però s´ha vist que existeixen dues formes que tenen dues estructures tridimensionals

diferenciades. La forma normal, anomenada PrP^{c} (c de cel·lular), i la forma patògena, anomenada PrP^{sc} (sc d´*scrapie*, nom provinent de la malaltia en ovelles, que fou la primera en estudiar-se).

Com es pot passar de la forma normal a la forma patogènica? En el cas de les malalties priòniques d´origen hereditari, simplement amb una mutació, és suficient perquè es produeixi el canvi d´estructura (és una proteïna força inestable). En el cas de les no hereditaries, només amb la interacció d´una forma patogènica (com en el cas de les vaques boges) amb la proteïna cel·lular normal és suficient perquè es produeixi una conversió lenta però exponencial de totes les proteïnes del cos a la forma patogènica. Aquesta forma té una propietat crucial; el cos no la pot eliminar i s´acumula a les neurones, fins que aquestes literalment exploten (degut a l´osmosi), deixant el cervell com una esponja. És per això que també s´anomenen malalties espongiformes transmissibles.

Vàrem pensar que seria interessant dissenyar un test diagnòstic de totes aquelles mutacions més freqüents que produeixen aquest canvi de conformació de la proteïna, i que per tant produeixen patogenicitat. Ens van interessar també els canvis a nivell d´ADN que tinguessin implicacions a nivell de susceptibilitat a les malalties. S´havia descrit en la literatura que, segons la variant o al·lel que hi hagués en les posicions 129 i 219 de la proteïna priònica, es tenia més o menys risc de desenvolupar la malaltia. Així, s´havia vist que en el codó 129, ser homozigot (i per tant, tenir en els dos cromosomes, un provenint del pare i l´altre de la mare, la mateixa variant) incrementava la probabilitat de desenvolupar la malaltia. En el cas de la posició 219, era el

contrari, ja que el fet de tenir la variant lisina (i no l´àcid glutàmic), protegia d´adquirir aquestes malalties priòniques. Però si alguna de les posicions 102, 178 o 200 del gen, estigués mutada, indubtablement es desenvoluparia una malaltia priònica familiar. En el cas de la mutació en la posició 102, es desenvolupa GSS; en la 178, FFI o CJD familiar -depenent de la variant existent en la posició 129-, i si la mutació afecta la posició 200, es manifesta CJD familiar. Així, al mateix temps i en un únic test, podíem veure directament totes les posicions del gen que ens interessaven: les més importants, les que estan relacionades amb malalties genètiques d´origen priònic, i també les posicions que tot i que no impliquen malalties hereditàries, tenen un pes important en les malalties priòniques adquirides i en les espontànies. Per tant, de forma ràpida i eficaç podíem obtenir tota la informació genètica, ja que mitjançant un codi establert que vàrem dissenyar podíem saber directament si el pacient desenvoluparia una malaltia priònica familiar o bé, si tenia més o menys probabilitat de desenvolupar alguna altra de les dues formes de malalties priòniques.

Paral·lelament, en el mateix estudi, vàrem aplicar aquest mètode diagnòstic a diferents mostres representatives de la variació genètica humana, utilitzant ADN de varies poblacions, tant europees, africanes, asiàtiques, ameríndies, i també d´Oceania. El resultat fou sorprenent, perquè les variants trobades en les posicions 129 i 219 mostren que algunes poblacions tenen molt més risc de desenvolupar CJD. Això és degut a que alguns grups humans tenen amb una freqüència més alta unes variants i en altres poblacions aquestes variants són

en una freqüència molt més baixa o són fins i tot inexistents. Així, les poblacions asiàtiques, *a priori*, tindrien més risc a patir aquests tipus de malalties, essent les poblacions ameríndies i les europees, les menys susceptibles. Aquesta interpretació fou possible gràcies a que es coneix per estudis previs que els heterozigots (és a dir, els que tenen en un cromosoma l´al·lel metionina i en l´altre cromosoma, l´ al·lel valina) per la posició 129, són els menys proclius a tenir malalties d´origen priònic (tal com hem vist en el cas del kuru a Nova Guinea). Tot i així, vèiem com a la posició 219, existia un al·lel, la lisina, que tenia un efecte protector, i aquest es trobava única i exclusivament en poblacions asiàtiques, de l´Orient Mitjà i poblacions d´Oceania. Curiosament doncs, trobem que hi hauria un fenomen compensatori parcial (ja que l´al·lel lisina es troba en freqüències molt baixes), ja que són les mateixes poblacions que tenen un alt risc de desenvolupar malalties priòniques degut a la posició 129, les que tenen també aquest al·lel protector en la posició 219. Tot i així, en general, veiem que en les poblacions d´Àsia, i més especialment, les del centre i est del continent, el risc d´adquirir malalties priòniques és més alt, tot i que per sort, al no utilitzar el mateix protocol en l´el·laboració de farines que el Regne Unit, no han patit el fenomen de les vaques boges, ja que sinó possiblement els resultats haurien estat molt pitjors per a l´espècie humana.

Aquestes diferències en les freqüències de les variants del gen PRNP entre les poblacions són producte de pressions selectives importants. Com veiem, la selecció natural pot tenir un gran efecte en molt poques generacions. Mirem sinó què va passar a Nova Guinea amb el kuru.

La proteïna priònica i els primats

Després de la importància observada pel codó 129 i altres posicions en el gen ens vam plantejar la possibilitat d´estudiar l´espècie més pròxima a la humana, el ximpanzé. Ens vam formular un seguit de preguntes: tenia també tanta importància el codó 129? existien també dues variants (metionina i valina) en aquesta posició? i si existien, tenien un avantatge selectiu els heterozigots?

A partir d´ una de les mostres més grans de ximpanzés mai estudiades, ens vàrem decidir a analitzar la variabilitat en el gen PRNP dins aquesta espècie, i per fer-ho vàrem seqüenciar-lo (és a dir, resoldre la combinació de lletres A, T, C i G que conformen aquest gen) per a totes les mostres. També vam seqüenciar el gen en una mostra més petita, però igualment informativa, de goril·les.

Els resultats obtinguts indicaven que el polimorfisme (l´existència de més d´una variant en una posició determinada), en aquest cas la posició 129, era exclusiva de l´espècie humana. Tant en ximpanzés com en goril·les observàvem només una variant, l´al·lel metionina, i forçosament, tots els ximpanzés i goril·les eren homozigots per a la posició 129; *a priori*, doncs, tindrien una major susceptibilitat a desenvolupar malalties priòniques que els humans. Curiosament, cal subratllar que en el codó 219, on hi havia l´al·lel lisina que protegia de la malaltia, no s´ha trobat present en cap de les mostres, i que per tant, en els nostres primats germans, aquesta posició era monomòrfica, i ho era per l´al·lel que no tenia cap efecte protector. Això ens suggeria que la pressió en aquest gen havia afavorit el sorgiment de l´altra

variant i que aquesta havia estat seleccionada en l´espècie humana per protegir-nos de fenòmens similars als de la malaltia del kuru al llarg de la nostra història.

Així doncs, el gen de la proteïna priònica és un exemple clar i molt suggeridor de l´efecte que la selecció natural pot exercir sobre el genoma, i més encara, que el canvi d´una base (que produeix que es codifiqui o metionina o bé valina en el codó 129) pot ser determinant per sortir victoriós de períodes caníbals de l´espècie humana. Una única base, de les 3 mil milions de bases que tenim en el genoma.

INTRODUCTION

1 PRION DISEASES

1.1 Introduction

Creutzfeldt–Jakob disease (CJD), Kuru, Gerstmann–Sträussler syndrome (GSS) and fatal familial insomnia (FFI) in humans, as well as scrapie and bovine spongiform encephalopathy (BSE) in animals, are fatal neurodegenerative diseases that form part of the group of transmissible spongiform encephalopathies (TSEs), or prion-related diseases. TSE are characterized by vacuolation of neurons, astroglyosis and the accumulation of the abnormal, protease-resistant prion protein in the central nervous system (CNS) (Ironside and Bell 1997). In animals, the most common TSE disease is *scrapie* (Kimberlin 1986), a disorder affecting sheeps and goats that was first recognized in 1730. The name is derived from the observation that affected animals rub themselves against the fences of their pens to stay upright, presumably reflecting the manifestation of ataxia. Eventually, the animals become unsteady on their feet, uninterested in their surroundings, and stop feeding. In humans, the most common TSE is CJD, with a worldwide incidence of 0.5–1.5 new cases per one million people each year (Johnson and Gibbs 1998). Patients are usually between 50 and 75 years old, and typical clinical features include a rapidly progressive dementia, myoclonus and a characteristic electroencephalographic pattern. Traditionally, three different forms of CJD have been identified: sporadic (85% of cases), familial (10%) and iatrogenic (~5%). Variant CJD (vCJD) is a new disease, which was first described in 1996 (Will et al. 1996). In contrast to typical cases of sporadic CJD, this variant form affects young

patients (27 years old on average) and lasts a relatively longtime (14 months versus 4.5 in sporadic CJD). Despite the fact that TSEs are relatively rare diseases, they have gained significant attention from both the scientific community and society in general. There are two major reasons for the high impact of prion research. First, the scare originated by the appearance of vCJD, and evidence linking it to the exposure of humans to the BSE agent, most plausibly owing to dietary contamination by affected bovine CNS tissue. Because of insufficient information available regarding the incubation time and the actual level of exposure to the contaminated material, it is impossible to make any well-founded predictions about the future number of vCJD cases. This uncertainty makes the situation even more frightening. Second, the field is overwhelmed by several unprecedented scientific findings, which have directly confronted some of the most popular dogmas in biology. These heretical findings include:

(1) disease transmission by an 'infectious protein' in the absence of nucleic acid;

(2) the folding of a protein in several different conformations with distinct biological properties; and

(3) the transmission of biological information by the replication of protein conformation.

These findings have given support for an entirely novel disease mechanism, involving transmission of *protein misfolding* (Prusiner 1998). It is important to note that despite the strong evidence supporting this conclusion, it still remains

a hypothesis that needs further validation (Chesebro 1998) (Mestel 1996) (Dormont 1999) (Carp et al. 1994).

Disease propagation without nucleic acid: an infectious protein

In 1967, the radiologist Alper and colleagues reported that the agent responsible for scrapie was extremely resistant to treatments that normally destroy nucleic acids, such as exposure to UV and ionizing radiation (Alper et al. 1967). This finding suggested the infectious material could in fact lack nucleic acid. This finding represented the first clash with the biological dogma that conventional pathogens are unable to replicate in the absence of nucleic acids. These results, together with their previous finding that the minimum molecular weight to maintain infectivity was too small ($\sim 2 \times 10^5$) to possibly be a virus or any other known type of infectious agent (Alper et al. 1966), led J.S. Griffith to propose that, uniquely, the material responsible for the disease transmission was a protein that has the surprising ability to replicate itself in the body (Griffith 1967). This was the beginning of the so-called 'protein-only' hypothesis of TSE propagation. In the past 30 years, many research groups worldwide have unsuccessfully attempted to find a virus, or at least a nucleic acid, associated with the disease (Prusiner 1998). Since the 1980s, research supporting the *protein-only hypothesis* of TSE transmissibility has been pionnered by Stanley Prusiner's group – who also coined the name 'prion'. A crucial finding for the understanding of the nature of this novel agent was the isolation of the prion protein (PrP) from the infectious material (Bolton et al. 1982). It was shown by immuno- and bio-chemical methods that PrP and

scrapie infectivity copurified, and that the concentration of the protein is proportional to the infectivity titer (Gabizon et al. 1988). Further purification of the protein to a level at which no other component is detectable, retains infectivity. Moreover, infectivity was convincingly neutralized by agents that destroy protein structure and, more importantly, by anti-PrP antibodies (Gabizon et al. 1988). Purification of PrP made it possible to obtain N-terminal amino acid sequence and, ultimately, cDNAclones encoding PrP protein. PrP mRNA proved to be the product of a single host gene, which was present in the brain of uninfected animals, and was constitutively expressed by many cell types (Prusiner 1991b). Genetic studies have shown that most, if not all, of the familial cases of TSE are linked to mutations in the *PRNP gene* (Aguzzi and Brandner 1999), and that mice expressing *PRNP gene*s with point mutations linked to GSS spontaneously develop neurological dysfunction, spongiform brain degeneration and astrocytic gliosis (Hsiao et al. 1990). However, it is not clear whether the cerebral damage observed is due to the presence of the mutation, or to overexpression of the *PRNP gene* (Manson et al. 1999). Another piece of evidence to support the prion hypothesis came from Charles Weissman's group, who showed that mice devoid of the *PRNP gene* are resistant to scrapie infection, neither developing symptoms of scrapie nor allowing propagation of the infectious agent (Bueler et al. 1993). This finding also gave some clues regarding the mechanism of prion replication, demonstrating that PrP host was required for disease propagation. Since the discovery of nucleic acids, the reproduction ability and heritage are considered to be determined and ruled only by the presence of specific sequences of DNA or RNA. Nevertheless, in the

early 19th century, it was commonly thought that the genetic information required for inheritance was carried by the conformational complexity that only proteins possess. The discovery of DNA came as a surprise to many, and was not generally accepted until the early 1950s. Until now, it was believed that the `replication´ of biological information in nature – from the simplest organisms (virino, virus and bacteria) to the most complex mammals – requires replication of nucleic acids (DNA or RNA), the carriers of hereditary information. Prion diseases take a step back to the pre-DNA era and could bring a new concept to the laws of inheritance – 'conformational heritage'.

One sequence, two conformations

The human prion protein is a product of a single gene located on the short arm of chromosome 20 (Prusiner 1991a) (Fig. 0).

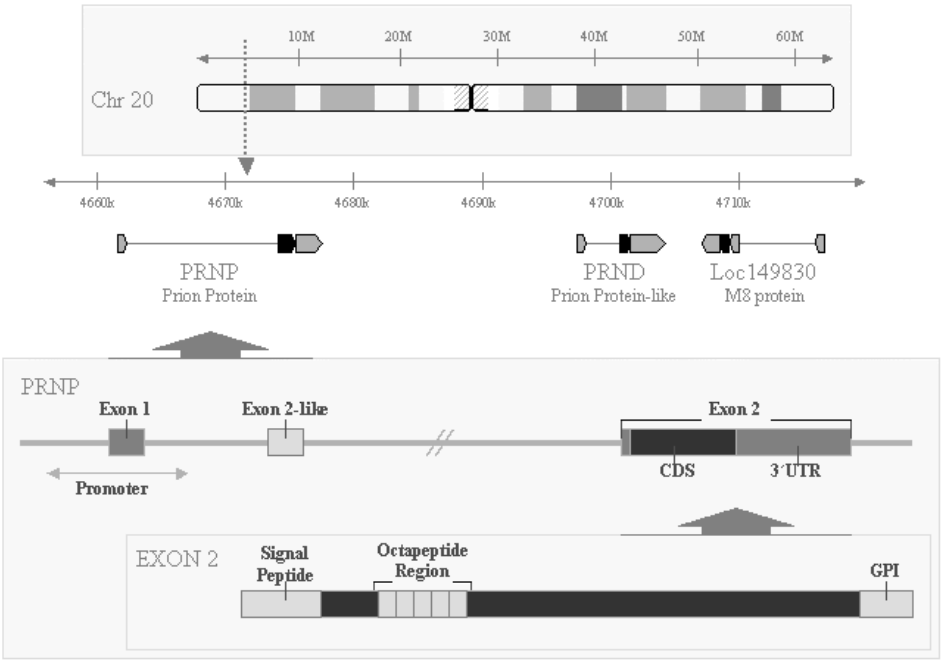


Figure0. *PRNP* gene

The mRNAproduct is 253 amino-acids long, contains five amino-terminal octapeptide repeats, two glycosylated sites and one disulfide bridge. In addition, there are two signal sequences in the amino- and carboxy-terminal ends, which are removed during processing, and a glycosylphosphatidylinositol anchor (GPI) that attaches the protein to the outer surface of the cell membrane (Harris 1999). The *PRNP gene* is constitutively expressed in the brain and other tissues of healthy animals and humans. To differentiate the normal protein from the one present in the infectious material, the normal protein has been termed PrPC (for cellular prion protein) and the one present in the infectious material has been referred as PrPSc (for scrapie-associated prion protein). Both proteins are identical in size and N-terminal sequence, and have a similar glycosylation pattern (Prusiner 1998). Strikingly, when limited protease digestion was carried out, PrPC was completely degraded, whereas PrPSc was only partially cleaved, removing a fragment from the amino terminus to produce the same molecular weight (27–30 kDa) as the protein previously purified from scrapie-infected brain (Bolton et al. 1982). Thus, PrP appears to have at least two distinct forms: a protease-sensitive state, found ubiquitously, and a protease-resistant state, found in the disease setting (Fig. 1). Although PrPC is a soluble protein, the PrPSc form is stubbornly insoluble, and forms aggregates in infected brain parenchyma (Fig. 1). Whether neurodegeneration is induced by soluble or aggregated PrPSc is still unclear.

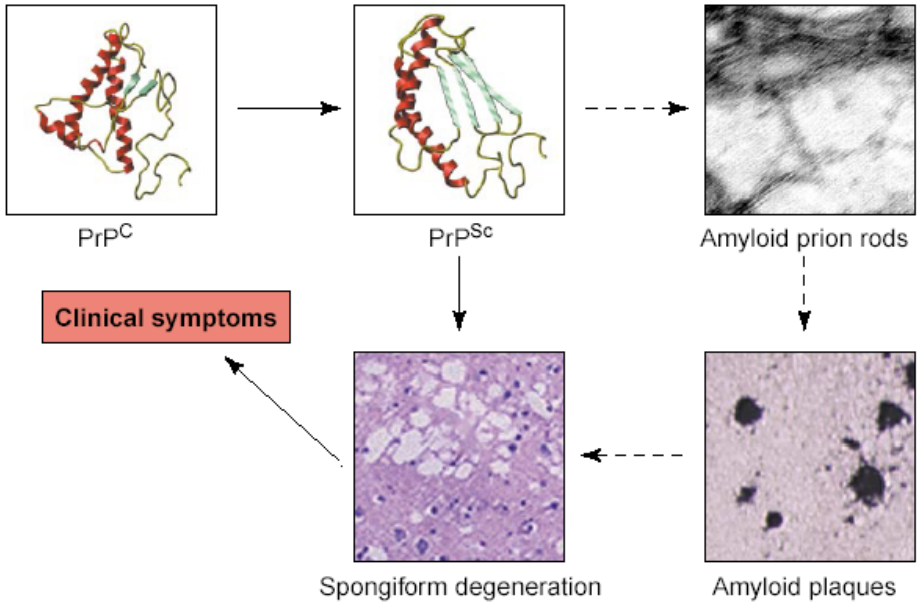


Figure1. A model for the progression of transmissible spongiform encephalopathty (TSE) pathogenesis. The hallmark event in the disease is the misfolding of the normal PrPc to form the B-sheet-rich PrPSc. The misfolded protein has a high tendency to aggregate, forming amyloid prion rods, which might be the precursor of cerebral amyloid plaques seen in some TSE cases. From (Soto and Saborio 2001).

Although PrPSc has a high-tendency to aggregate *in vitro* forming *amyloid* prion rods, there is only extensive accumulation of *amyloid* in the brain in a small percentage of TSE cases (Prusiner 1998). In any case, the differences in protease resistance and solubility of the two PrP isoforms have been very useful for distinguishing them biochemically. Proteolytic resistance might be a key factor in enabling survival during the digestion process, and allowing its absorption by the gastrointestinal tract. Given that the primary structures of PrPC and PrPSc are identical, the process whereby the normal state of PrP protein is converted to the infection-associated form seemed likely to involve post-translational modifications. However, extensive biochemical characterization has failed to find any covalent differences between PrPC and

PrPSc (Stahl et al. 1993). By contrast, biophysical studies have demonstrated a dramatic conformational difference in the PrP forms (Pan et al. 1993). Techniques that allow the determination of secondary structure have shown that the α-helical content of PrPC is 40%, with little or no *β-sheet* content, whereas PrPSc contains 50% *β-sheet* and only 20% α-helix. The recently reported three-dimensional structure of PrPC from a variety of sources has allowed a direct determination of its secondary structure content, and the results confirm the secondary structure previously described (Riek et al. 1998).

The existence of two stable conformations for a polypeptide with the same amino acid sequence clashes with the general conception that the primary structure of a protein determines its unique three-dimensional structure. Anfinsen's classic denaturation–renaturation experiments of ribonuclease in the 1960s led to the proposal that the amino acid sequence of a protein contains the information required for folding into a unique native three-dimensional structure in a given environment (Anfinsen 1973). Since then, many other proteins have been refolded into their active form, and although the hypothesis has been challenged by the identification of molecular *chaperone proteins* which are able to alter protein folding, it remains a central dogma of structural biology. If the concept that a protein can adopt more than one folded conformation is heretical, the recent evidence of multiple conformations with different biological properties within PrPSc molecules (Caughey et al. 1998), (Safar et al. 1998) is even more surprising. A typical feature of spongiform encephalopathies is the existence of strain diversity (Kascsak et al. 1991).

Different prion strains within a single host species can be distinguished by reproducible differences in incubation period, clinical signs, distribution of neuropathological lesions and pattern of protease-resistant PrPSc. The existence of multiple prion strains has been the most common argument used against the *protein-only hypothesis* of prion transmissibility, as the basis for strain diversity in conventional pathogens is determined by variations in nucleic acid. However, in the recent past, evidence has accumulated to conclude that prion strains are the result of the existence of multiple PrPSc conformations (Kascsak et al. 1991) (Safar et al. 1998) (Bessen and Marsh 1994) (Aucouturier et al. 1999). These different conformations faithfully replicate in the body at the expense of normal PrPC. The different phenotypes, induced by diverse prion strains, probably result from distinct interactions with cerebral receptors, variable degrees of toxicity or different aggregation capabilities. Several research groups have shown that PrPSc molecules isolated from different strains in the same animal species exhibit clearly distinct conformations, which can be detected even with low resolution techniques to monitor secondary structure (Caughey et al. 1998) (Safar et al. 1998). These findings suggest that the three-dimensional structure of the protein associated with diverse strains must be quite different.

Replication of infectious agent by transmission of protein conformation

The notion that endogenous PrPC is involved in the development of infection is supported by experiments in which the endogenous *PRNP gene* was inactivated

by targeted deletion. PrP knock-out mice were both resistant to prion disease and unable to generate new infectious particles (Bueler et al. 1993). In addition, it is clear that during the time between the inoculation with the infectious protein and the appearance of the clinical symptoms, there is a dramatic increase in the amount of PrPSc ability to replicate itself at the expense of the host PrPC by transferring its toxic conformation to the normal protein. Prion replication is thought to occur when PrPSc in the infecting inoculum interacts specifically with host PrPC, catalyzing its structural conversion to the pathogenic form of the protein (Cohen and Prusiner 1998).

A physical association between the two isoforms during the infectious process is suggested by the primary sequence specificity in prion transmission, and by the reported *in vitro* generation of PrPSc-like molecules by mixing purified PrPC with PrPSc (Kocisko et al. 1994; Horiuchi and Caughey 1999). However, the exact mechanism underlying the conversion is not known. Investigations with chimeric transgenes showed that PrPC and PrPSc are likely to interact within a central domain delimited by codons 96 and 169 (Telling et al. 1995). These findings were supported by studies using several synthetic PrP peptides, which show that peptides between 109 and 141 reproduce some of PrP conformational flexibility *in vitro* (De Gioia et al. 1994), and also bind to PrPC and compete with PrPSc interaction (Chabry et al. 1998). More detailed biochemical studies should result in a further assessment of the size of the interactive region, as well as the definition of crucial residues. From molecular genetic studies (Telling et al. 1995), coupled with analysis of the requirements

for PrP conversion *in vitro (Saborio et al. 1999)*, the existence of a chaperone-like protein, provisionally called protein X, which facilitates the conversion of PrPC to PrPSc has been postulated (Cohen and Prusiner 1998). However, the nature of this factor remains unknown.

Assuming the *protein-only hypothesis* of prion propagation, the crucial step is the *conformational transmission* of the pathological PrPSc to the host PrPC. The precise molecular mechanism of PrPC to PrPSc conversion is not well understood. At least two hypotheses have been proposed (Fig2a and 2b):

(1) The 'template-assisted conversion' model (Cohen and Prusiner 1998), which postulates that PrPSc is thermodynamically more stable than PrPC, but that there is a kinetic barrier to reach stability. PrPC exists in equilibrium with a transient conformational intermediate (named PrP*), which, after interaction with a cellular chaperone (protein X), is able to heterodimerize with PrPSc. Spontaneously, this heterodimer is converted into a PrPSc homodimer, consisting of the old and newly formed PrPSc molecules (Fig. 2a). The homodimer can dissociate to form two templates, each of which is able to induce further conversion and thus generating an exponential growth of PrPSc concentration.

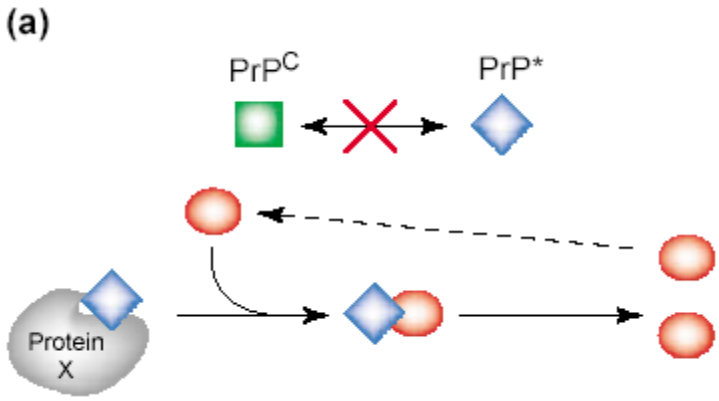


Figure 2a. Theoretical models for PrPC to PrPSc conversion. In the template-assisted conversion model, the key step is the formation of the intermediate state (PrP*), which, upon binding to a molecular chaperone (protein X), is capable of interacting with PrPSc, which acts as a template for the conversion. From (Soto and Saborio 2001).

(2) The 'nucleation–polymerization' model (Harper and Lansbury 1997) (Brown et al. 1991), which proposes the coexistence of PrPC and PrPSc in a thermodynamic equilibrium in solution. PrPSc monomer is unstable and becomes stabilized upon aggregation with other PrPSc molecules. PrPSc aggregates promote the conversion of PrPC by binding to the monomeric PrPSc and displacing the equilibrium toward the formation of the pathological conformer. In this model the infectious agent is a multimeric, highly ordered aggregate of PrPSc and the rate-limiting step is the formation of a nucleus that acts as a seed for further stabilization of PrPSc (Fig. 2b). Regardless of the molecular mechanism of PrP conversion, it seems clear that PrPSc is able to transfer its pathological conformation to PrPC, either by template mimicking or by stabilization upon aggregation. The transmission of phenotypic information through the progressive replication of protein conformation represents something

like a second heritable code, whereby a dominant protein conformation is transferred to the recessive form, changing its functional properties.

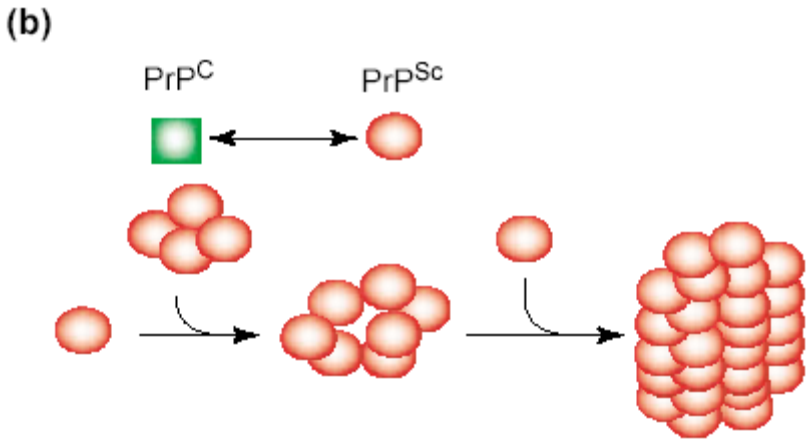


Figure 2b. Theoretical models for PrPC to PrPSc conversion. In the nucleation–polymerization model, the key step is the formation of a small PrPSc oligomer that acts as a seed for further stabilization of monomeric PrPSc, displacing the equilibrium towards formation of the pathological protein. From (Soto and Saborio 2001).

The first records of transmissible spongiform encephalopathies: scrapie

This term of scrapie is derived from a Scottish term, describing the tendency of afflicted animals to scrape their fleece against trees and bushes. But other words existed before in France (*la tremblante*, trembling) or in other countries such as Germany (*Gnubberkrankheit*, itching disease or *Traberkrankheit*, trotting)(Collinge et al., 1997). The first records appeared early in the 19th century, and it is known that this disease has been endemic in British sheep since 18th century, together with other European countries, such as Iceland (Thorgeirsdottir et al. 1999), where it appeared later.

It was when the transmissiblility of scrapie to goats was proved to be so efficient (100%), that the research community was convinced it was an infectious disease, and was thought for many years to be caused by a slow virus. It was at the end of last century when the scientifics generally accepted the protein only hypothesis.

Bovine spongiform encephalopathy (BSE)

Bovine spongiform encephalopathy (BSE) or "mad cow disease" appears to have originated from scrapie, an endemic spongiform encephalopathy of sheep and goats that has been identified in Europe since the mid-18th century (Brown and Bradley 1998). It has since spread to most sheep-breeding countries and is widespread in the United Kingdom (UK), where until 1988 the rendered carcasses of livestock (including sheep) were fed to ruminants and other animals as a protein-rich nutritional supplement. During rendering, carcasses from which all consumable parts had been removed were milled and then decomposed in large vats by boiling at atmospheric or higher pressures, producing an aqueous slurry of protein under a layer of fat (tallow). After the fat was removed, the slurry was desiccated into a meat and bone meal product that was packaged by the animal food industry and distributed to owners of livestock and other captive animals (e.g., zoo and laboratory animals, breeding species, pets). Although elements of the ensuing story are still disputed (including its origin from scrapie, rather than from unrecognized endemic BSE), it appears likely that changes in the rendering process that had taken place

around 1980 allowed the etiologic agent in infected carcasses to survive, contaminate the protein supplement, and infect cattle. Cattle carcasses and carcass wastes were then recycled through the rendering plants, increasing the levels of the now cattle-adapted pathogen in the protein supplement and eventually causing a full-scale BSE epidemic (Wells et al. 1987; Brown 1997; Collee and Bradley 1997b; Collee and Bradley 1997a). Recognition of this source of infection has led to a series of countermeasures taken by the UK and other countries to break the cycle of cattle reinfection, restrict the geographic spread of disease, and eliminate potential sources of new infections. Probably the single most important measure in the UK was the imposition in 1988 of a ruminant protein feed ban that by 1992 began to bring the epidemic under control. However, the loss of nearly 200,000 diseased cattle, followed by preemptive slaughter and destruction of nearly four and a half million asymptomatic cattle >30 months of age, has crippled the British livestock industry and also affected the tallow, gelatin, and pharmaceutical industries, all of which make bovine-derived products (Fig. 3).

BSE was not only found in UK. Cases have occurred in many other countries (like Spain or the USA) as a result of imported live animals or livestock food supplements. In some countries, including the UK, the incidence of new cases is decreasing, but in other countries–France, Portugal, Germany, Spain, and the Republic of Ireland–the incidence appears to still be increasing.

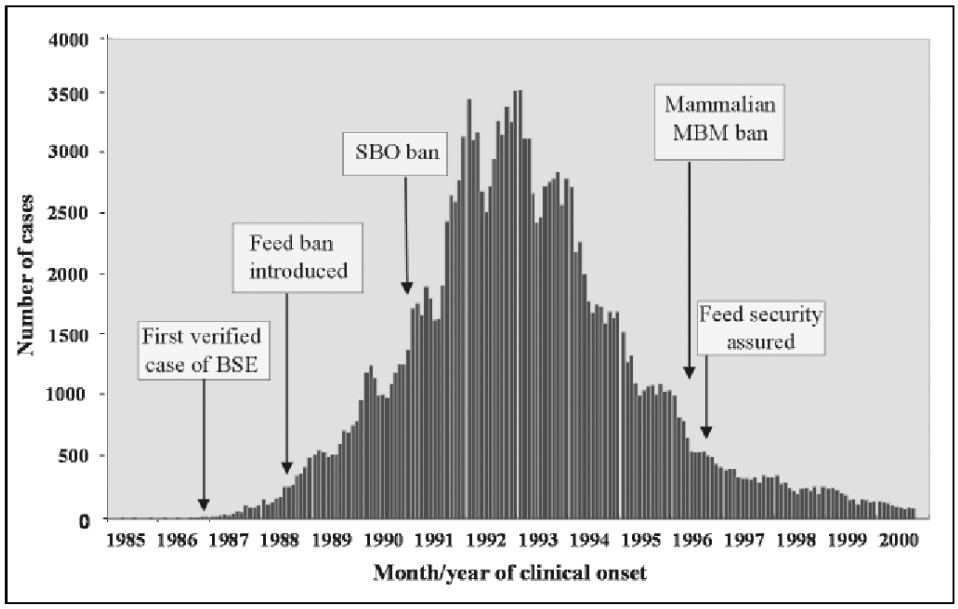


Figure 3. Time course of epidemic bovine spongiform encephalopathy in the United Kingdom, 1986-2000, with dates of major precautionary interventions. SBO= specified bovine offals (brain, spinal cord, thymus, tonsil, spleen, and intestines from cattle > 6 months of age) MBM= meat and bone meal (protein residue produced by rendering). From (Brown et al. 2001).

1.2. Classification of prion diseases

Human prion diseases are generally classified in the following way (Table 1):

Type	Clinical syndromes	Aetiology
Acquired	Variant CJD	BSE
	Kuru	Cannibalism
	Iatrogenic CJD	Inoculation
Sporadic	CJD	?Somatic PRNP mutation or spontaneous conversion of PrPc to PrPsc
	Atypical CJD	
Inherited	Familial CJD	Germline PRNP mutation
	GSS	
	FFI	
	Various atypical dementias	

Table 1. Modified from John Collinge et al. 1997

1.2.1. Acquired prion diseases

1.2.1.1. Variant Creutzfeldt Jakob disease (vCJD)

Within weeks of identification of the first case of BSE, concern was expressed about human risk (Holt and Phillips 1988) (Taylor 1989) (Kimberlin 1990) (Brown 1997) (Brown et al. 2001), and as the epidemic unfolded, a series of measures were taken to eradicate BSE and prevent potentially infected tissues from reaching the human food chain. A surveillance unit to monitor CJD was established in the UK in May 1990, and 3 years later, surveillance was extended to several other European countries, coordinated through the European Union. By this means it was hoped that any change in the epidemiology of CJD in the UK could be detected quickly and that the significance of the change could be assessed by comparison with the epidemiology of CJD in continental Europe. Concern was heightened by the discovery that some exotic zoo ungulates, as well as domestic and captive wild cats, were becoming infected (Jeffrey and Wells 1988) (Fleetwood and Furley 1990) (Aldhous 1990) (Young and Slocombe 2003) (Kirkwood et al. 1990) (Willoughby et al. 1992). The ungulates and domestic cats had also been fed diets supplemented by meat and bone meal, and the wild cats had been fed uncooked tissues, including cattle heads and spines. The possibility could therefore not be ignored that the disease might also cross the species barrier to humans from the consumption of beef or dairy products, or perhaps from occupational contact with cattle by ranchers, dairymen, or slaughterhouse workers. What most concerned about human infection was the presumption that BSE originated from scrapie, and scrapie

was not a human pathogen. Nevertheless, even those who considered human risk to be remote acknowledged that scrapie might unpredictably show an altered host range. Full comparative neuropathologic examination of both pre- and post-1980 cases of CJD in young persons was needed, along with *PRNP* gene sequence analysis of as many cases as possible. During February 1996, an additional case was referred to the Surveillance Unit with a clinical evolution similar to that of the previous seven patients, and neuropathologic examination of recent and historical cases confirmed that the recent cases were indeed distinctive. In particular, a morphologically unusual form of plaque was present in all cases: the florid or "daisy" plaque in which an *amyloid* core was surrounded by "petals" of spongiform change. As of March 1, despite the likelihood that this group of patients had a "variant" of CJD, it was still unclear whether mutations were involved and whether such a syndrome was also occurring outside the UK–both points essential to confirming the association of this variant disease with exposure to BSE. On March 4, genetic analysis was completed for six of the cases, and no pathogenic mutation was identified. These results effectively ruled out a genetic cause for the syndrome (although they did not rule out a genetic predisposition) and left the only remaining uncertainty–the geographic distribution of the variant phenotype–to be resolved by the European CJD surveillance system. The answer came by March 20: none of the young CJD patients in other European countries had the clinical and neuropathologic features of the UK cases. In the preceding week, two more variant cases had been neuropathologically confirmed, and a report on the entire group of 10 cases concluded that an unrecognized variant of CJD

occurring only in persons <45 years of age was probably due to exposure to BSE (Will et al. 1996). This link has now been convincingly established in laboratory studies showing identical, distinctive biological and molecular biological features of the pathologic agent isolated from BSE-infected cattle and human cases of vCJD (Collinge et al. 1996) (Bruce et al. 1997) (Scott et al. 1999). The source of contamination appears to have been beef. However, muscle has never been reproducibly shown to contain the infectious agent in any form of spongiform encephalopathy, whatever the affected species, and thus, infection most probably resulted from beef products contaminated by nervous system tissue. Contamination could have occurred in any of the following ways: cerebral vascular emboli from cranial stunning instruments used to immobilize cattle before killing by exsanguination; contact of muscle with brain or spinal cord tissue by saws or other tools used during slaughter; inclusion of paraspinal ganglia in cuts of meat containing vertebral tissue (e.g., T-bone steaks); and perhaps most importantly, the presence of residual spinal cord and paraspinal ganglia tissue in the paste of "mechanically recovered meat" (a carcass compression extract) that could legally be added to cooked meat products such as meat pies, beef sausages, and various canned meat preparations. Measures have since been taken to eliminate these sources of potential contamination and limit the consequences of any contamination that may already have occurred.

Some of the features of this syndrome (vCJD) are reminiscent of kuru, in which behavioural changes and progressive ataxia predominate. The age at onset in

the initial cases reported ranged from 16 to 39 years and the clinical course was unusually prolonged (median 12 months). The EEG was atypical, without the pseudoperiodic pattern seen in most sporadic SJD cases. The neuropathological appearances were striking and consistent. There was widespread spongiform change, gliosis, and neuronal loss, most severe in the basal ganglia and thalamus, but the most remarkable feature was the abundant PrP *amyloid* plaques in the cerebral and cerebellar cortex. These consisted of kuru-like, florid (surrounded by spongiform vacuoles) and multicentric plaque types. The florid plaques, seen previously in scrapie, were a particularly unusual but highly consistent feature. (Collinge et al., 1997)

Although the amount of infectious tissue ingested must be a critical determinant for the transmission of BSE to humans in the form of vCJD, the human genotype at polymorphic codon 129 of the PRNP gene appears to play an important role in susceptibility to infection. The encoding alternatives, methionine (Met) and valine (Val), are distributed in the general Caucasian population in the approximate proportions of 50% Met/Val, 40% Met/Met, and 10% Val/Val. All 76 vCJD patients tested have been homozygous for methionine, and the apparently single infecting strain of BSE may not be able to replicate in any other human genotype. However, it is also possible that (as in the analogous oral infection of kuru and in peripheral iatrogenic CJD infections) heterozygotes are comparatively resistant to disease and become ill after longer incubation periods than those of homozygotes (Cervenakova et al. 1998) (Lee et al. 2001) (Huillard d'Aignaux et al. 1999) (Brown et al. 2000).

1.2.1.2. Kuru

Kuru reached epidemic proportions amongst a defined population living in the Okapa district of the eastern highlands of Papua New Guinea during the 1950s-60s (Fig. 4).

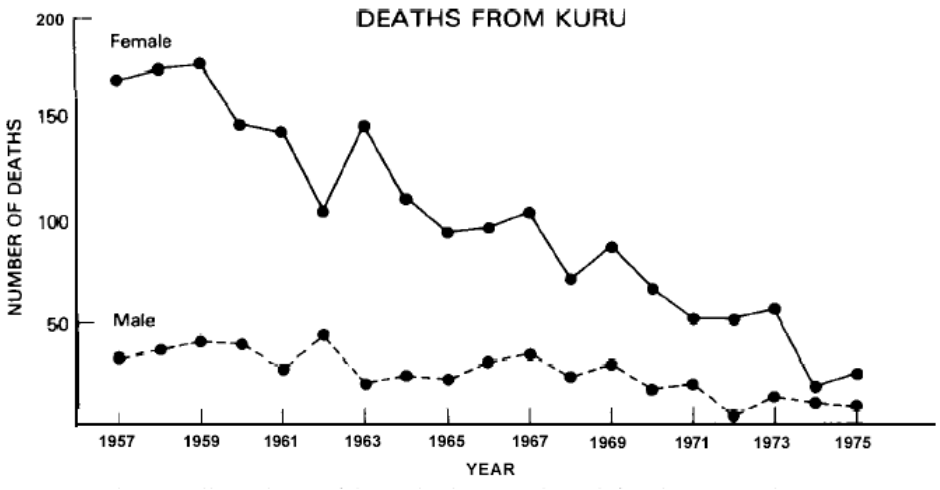


Figure 4. The overall incidence of kuru deaths in male and female patients by year since its discovery in 1957 through 1975. More thatn 2.500 patients died of kuru in this 17 year period of surveillance. The decline in incidence of the disease has followed the cessation of cannibalism, which ocurred between 1957 and 1962 in various villages. From Gajdusek 1976.

Kuru is characterized by cerebellar ataxia and a shivering-like tremor that progresses to complete motor incapacity and death in less than one year from onset. It was confined to a number of adjacent valleys in the mountainous interior of New Guinea and occurred in 160 villages with a total population of just over 35,000 (Figs. 5-6). *Kuru* means shivering or trembling in the Fore language.

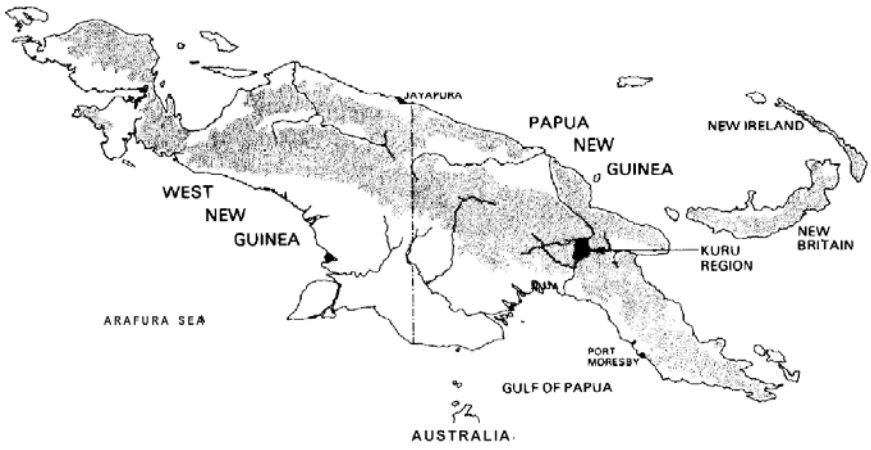


Figure 5. Papua New Guinea Map. The kuru region is indicated. From Gajdusek 1976.

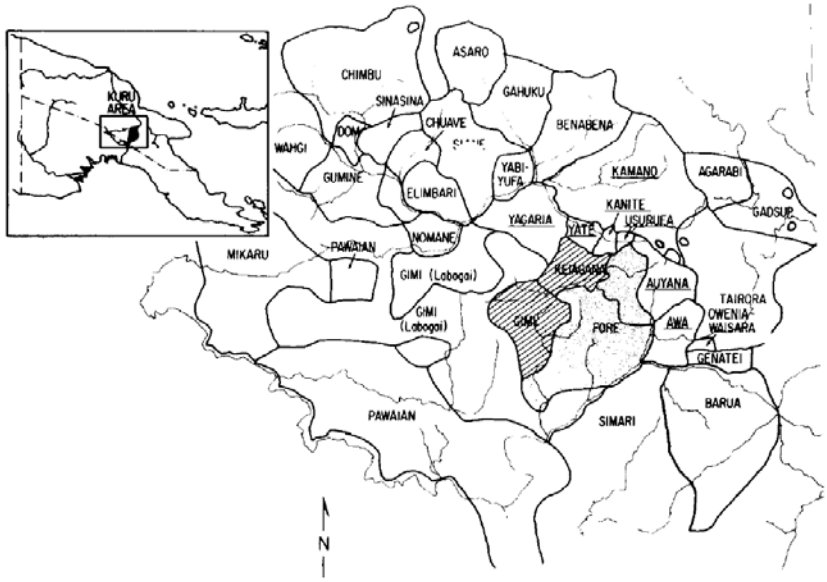


Figure 6 The kuru region in the Eastern Highlands Province of Papua New Guinea showing the cultural and linguistic groups in and surrounding the kuru affected populations. *Inset,* upper left: Eastern half of the island of New Guinea showing, in rectangle, area included in the map of larger scale. . From Gajdusek 1976.

In the Fore culture and linguistic group, among whom over 80 % of the cases occurred, it had a yearly incidence rate and prevalence ratio of about 1 % of the population. During the early years of investigation, after the first description by Gajdusek and Zigas in 1957 (Gajdusek and Zigas 1957), it was found to affect all ages beyond infants and toddlers; it was common in male and female

children and in adult females, but rare in adult males. This marked excess of deaths of adult females over males has led to a male-to-female ratio of over 3:1 in some villages, and of 2:1 for the whole South Fore group. Kuru has been disappearing gradually and the disease is no longer seen in either children or adolescents. This change in occurrence of kuru appears to result from the cessation of the practice of ritual cannibalism (which occurred between 1957 and 1962 in various villages) as a rite of mourning and respect for dead kinsmen, with its resulting conjunctival, nasal, and skin contamination with highly infectious brain tissue mostly among women and small children.

Similar to other the TSEs, kuru had a long incubation period; it was years or even decades before an infected person showed symptoms. Because kuru mainly affected the cerebellum, which is responsible for coordination, the usual first symptoms were an unsteady gait, tremors, and slurred speech. Unlike most of the other TSEs, dementia was either minimal or absent. Mood changes were often present. Eventually, individuals became unable to stand or eat, and they died in a comatose state from 6 to 12 months after the first appearance of symptoms.

The American physician D. Carleton Gajdusek established the infectious nature of the disease by injecting samples of brain tissue from subjects who had died from kuru into the brains of chimpanzees; the primates eventually developed and succumbed to the disease.

It is hypothesized that kuru originated from consumption of an individual with sporadic CJD (Alpers and Rail 1971), a disease with a remarkably uniform

worldwide incidence of around 1 per million and a lifetime risk of around 1 in 50,000.

In a study by Cervenakova et al (Cervenakova et al. 1998), they found that in a total of 92 kuru patients, homozygosity at codon 129 (particularly for methionine) was associated with an earlier age at onset and a shorter duration of illness than was heterozygosity, but other clinical characteristics were similar for all genotypes. (Fig. 7)

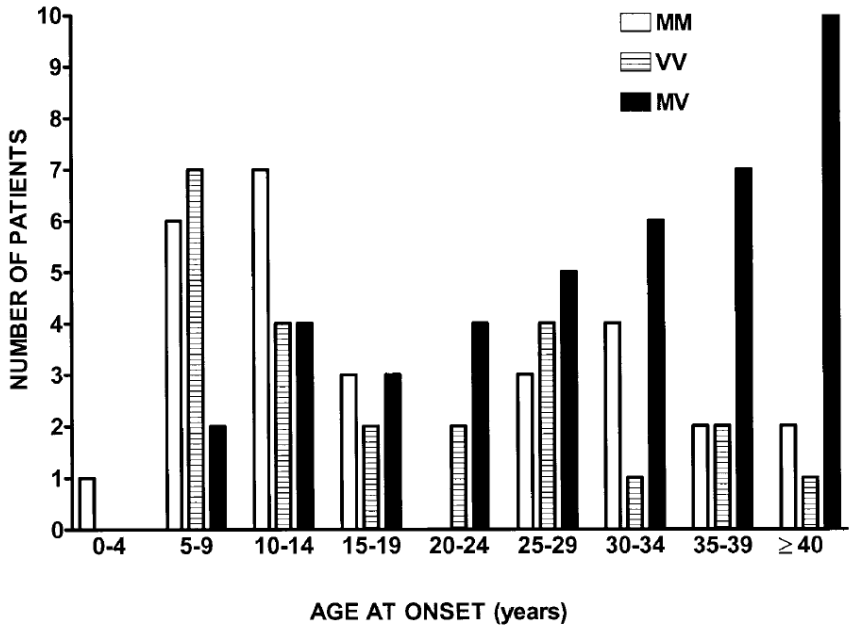


Figure 7. Distribution of PRNP codon 129 genotypes according to age at onset of illness in 92 kuru patients. From (Cervenakova et al. 1998).

In another study by Mead et al (Mead et al. 2003), blood was obtained from Fore women aged 50 years or above. They had a history of multiple exposure to mortuary feasts. Twenty-three out of 30 women over the age of 50 were heterozygotes at codon 129, a finding that is significant (Fisher's exact test, $P = 0.01$) compared with the genotypes of the unexposed Fore population, which

are in Hardy-Weinberg equilibrium (*n* =140). Two large samples of elderly Europeans also displayed Hardy-Weinberg equilibrium, which suggested that this age effect was local to the Fore. The age of onset of kuru in homozygotes of either allele at *PRNP* codon 129 has been estimated around 19 years, but over 30 years for heterozygotes (Cervenakova et al. 1998). Thus the marked survival advantage for codon 129 heterozygotes provided a powerful basis for selection pressure in the Fore.

1.2.1.3. Iatrogenic Creutzfeldt-Jakob disease (iCJD)

Iatrogenic prion disease is usually referred to as CJD although cases arising as a result of peripheral (rather than intracerebral) inoculation with prions usually have a clinical picture more reminiscent of kuru, with a prominent cerebellar syndrome.

Transmission of CJD from case to case has occurred by a number of routes involving accidental inoculation with human prions as a result of medical procedures. Such iatrogenic routes include the use of inadequately sterilized neurosurgical instruments, dura mater and corneal grafting, and the use of human cadaveric pituitary-derived growth hormone or gonadotrophin.

Head et al. (Head et al. 2003) found that presumptive centrifugal spread of PrP from the brain through the optic nerve occurred in both sporadic and variant CJD. Given that routine decontamination might not remove PrP from surgical instruments, the authors proposed that a careful risk assessment be made of possible iatrogenic spread of sporadic and variant CJD after surgery on the retina or optic nerve.

Brown et al. (Brown et al. 1994) tested 15 cases of iatrogenic CJD that represented central infection (from dura mater or corneal homografts and stereotactic EEG electrodes), 11 cases peripherally infected (from native human growth hormone or gonadotropin), and 110 control individuals for the presence of mutations in *PRNP* gene. No patient or control had any of the known pathogenic point or insertional mutations found in the familial disease, but allelic homozygosity at codon 129 was present in all but 2 (92%) of the 26 patients, compared with 54 (50%) of the 110 controls (p less than 0.001). Pooled data from all identified and tested cases of iatrogenic disease yielded a worldwide total of 56 patients, of whom all but 4 were homozygous at codon 129 (p less than 0.001).

1.2.2. Sporadic prion disease

1.2.2.1. Creutzfeldt-Jakob disease (sCJD)

sCJD represents about 80-85% of the total cases of human prion diseases. Any familial mutation is associated with these sporadic cases and no evidence of any type of transmission neither environmental nor genetic or external inoculation has seen.

The overall annual incidence in most studies was around 0.5-1 case per million. There is no significant case clustering other than in familial clusters. Cases are distributed apparently at random with a frequency related only to local

population density. There is no evidence of an association with local scrapie prevalence. For instance, CJD is more common in Australia and New Zealand, which have been scrapie free for many years, than in the UK where scrapie is endemic.

The core clinical syndrome of CJD is of a rapidly progressive multifocal dementia usually coupled to myoclonus. The onset is usually in the 45-75 year age group with peak onset between 60-65. Around 70% of cases die in under six months.

Neuropathological confirmation of CJD is by demonstrated by spongiform change, neuronal loss, and astrocytosis. PrP *amyloid* plaques are usually not present in CJD although protease-resistant PrP, seen in all the currently recognized prion diseases, can be demonstrated by immunoblotting of brain homogenates.

Genetic susceptibility has been demonstrated in that most cases of classical CJD are homozygous with respect to 129 polymorphism of PrP. (Alperovitch et al. 1999).

A significant independent association was found between sCJD and a polymorphism upstream of *PRNP* exon 1 (position 1368) (Mead et al. 2001), in addition to the strong susceptibility conferred by codon 129. However, although their sample size was necessarily small, no association was found between these polymorphisms and vCJD or iatrogenic CJD. This observation can be

important but until now, no bigger study has been performed regarding this observation.

1.2.3. Inherited prion diseases

Inherited prion diseases are prion diseases caused by mutations in the PRNP gene (Fig. 8). A list of the mutations considered to be causative can be found in Table 2, or in the Mad Cow official Home Page (http://www.mad-cow.org/prion_point_mutations.html). There are two types of mutations, SNP (or single nucleotide polymorphisms), that affect only one nucleotide, or INDELS (insertions or deletions), where more than one nucleotide is affected. This last category is referred to the octapeptide region, which is a minisatellite. Extra copies are sometimes referred to insertions, and loss of some copies are sometimes referred to deletions.

Figure 8. Human Prion protein gene.

Polymorphism		Mutation		
Silent	Influential	Point		Insertional
P68P	M129V	P102L	T188A	24bp
A117A	N171S?	P105L	T188K	48bp
G124G	E219K?	A117V	**E196K**	96bp
V161V	24bp deletion?	G131V	F198S	120bp
N173N*		I138M*	**E200K**	144bp
H177H		G142S*	D202N	168bp
T188T*		Y145s	**V203I**	192bp
D202D		Q160s	**R208H**	216bp
Q212Q		**D178N–129V**	**V210I**	
R228R		*D178N–129M*	**E211Q**	
S230S		**V180I**	Q212P	
		V180I + M232R	Q217R	
		T183A	**M232R**	
		H187R	M232T	
		T188R	P238S	

Table 2. Variations in the Human Prion Protein Gene coding region. Bold indicates CJD phenotype, underlined indicates GSS, itaics indicate FFI. Others are not categorised, as the published data are insufficient, or findings are unusual to the known disease subtypes. *Referred from Mad Cow Disease Home Page. From (Kovacs et al. 2002).

1.2.3.1. Familial Creutzfeldt-Jakob disease (fCJD)

This disorder is caused by mutations in the prion protein gene (PRNP). Sometimes is also termed inherited or genetic. This is because a low percentage of fCJD cases have familial history of prion diseases.

Masters et al. (Masters et al. 1979) found that about 15% of CJD cases are familial. From a study of 73 families, (Masters et al. 1981) concluded that 15% of cases of CJD have a family history consistent with autosomal dominant transmission. Onset of disease is significantly earlier in familial cases (Fig. 9). A maternal effect has not been found. Temporal and spatial separations between affected relatives suggested that incubation periods range at least from 1 to 4 decades. Affected sibs tend to die at the same age and not at the same time (Masters et al. 1981). In 4 families, CJD occurred in members related by marriage.(Bertoni et al. 1983) reported 7 affected persons in 3 generations of a large kindred. They pointed out that 3 of 4 patients studied in detail were first observed with supranuclear gaze paralysis, gait ataxia, and rapidly progressive dementia. Most of the affected persons were farmers.

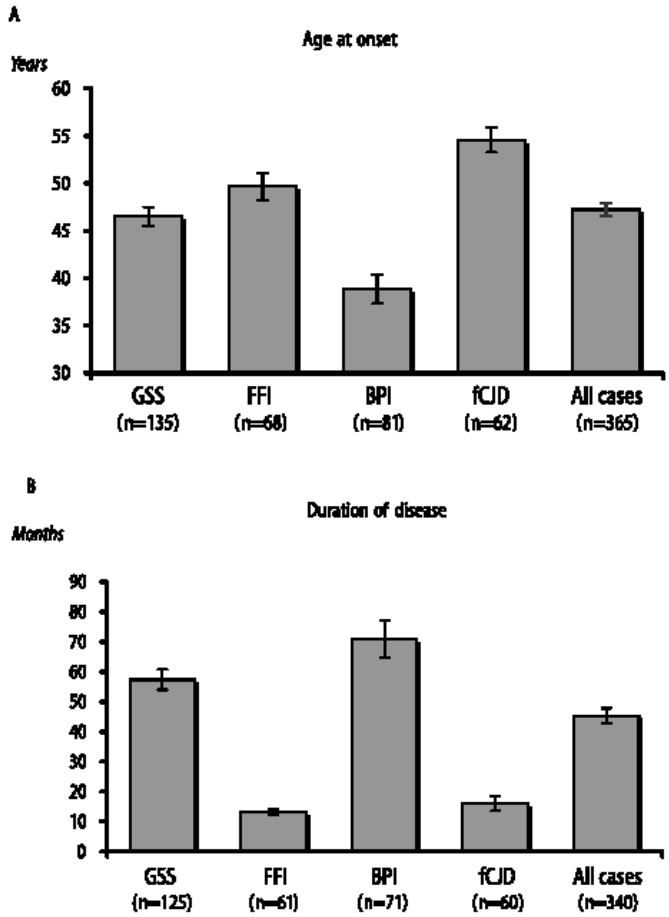


Figure 9. (A) Mean age at onset calculated in years in the different disease groups. (B) Mean duration of illness calculated in months in the different disease groups. FCJD familial Creutzfeldt-jakob disease; GSS Gerstmann-Straussler-Scheinker disease; BPI familal CJD caused by base pair insertion; FFI fatal familial insomnia. From (Kovacs et al. 2002).

(Brown et al. 1984) analyzed the characteristics of those 5 to 10% of patients who pursue a relatively long course (more than 2 years). Patients with prolonged course showed higher familial representation (30%), younger age of onset (average, 48 years), and lower frequency of myoclonus (79%) and periodic EEG activity (45%) than are found in series of unselected cases. The longest course was 13 years in a case proved by transmissibility. Of 225 transmitted cases, 15 (7%) had a prolonged course. The incubation period and

duration of illness after injection into primates bore no relation to the duration of illness in patients. Clinical differentiation from Alzheimer disease can be difficult in CJD patients with a long course.

In a consecutive series of 230 patients with neuropathologically verified CJD, (Brown et al. 1986) found that men and women were affected about equally with a mean age of onset of 61.5 years. Familial cases accounted for 4 to 8% of the series. Most of the early neurologic symptoms were cerebellar or visual. Extrapyramidal muscular rigidity, myoclonus, and characteristic periodic EEG complexes were observed comparatively late. The median duration of illness was 4 months and the mean was 7.6 months; 90% of patients died within a year of onset.

Creutzfeldt-Jakob disease occurs at an unusually high frequency in Chile (Masters et al. 1979). (Kahana et al. 1974) described an aggregation of cases among Libyan Jews, a finding that supports the viral or the genetic hypothesis or perhaps both. In a country-wide survey of CJD in Israel, (Zilber et al. 1991) diagnosed 114 cases, among them 49 Libyan-born, with onset of their disease during the years 1963-1987. After age adjustment, the mean annual incidence rate per million population was 43 among Libyan-born and 0.9 in the rest of the population. Among Jews born in Egypt and Tunisia, countries neighbouring Libya, the adjusted rates were higher than in the other Israelis (3.5 and 2.3 per million, respectively). Among Libyan Jews, there was no association between incidence rate of CJD and age at immigration, i.e., duration of exposure to a

hypothetical infectious factor in Libya. The percentage of familial cases among Libyan Jews (41 to 47%) is one of the highest known.

(Kahana et al. 1991) reported that the clinical presentation and evolution of the disease were very similar in patients born in Libya and others without Libyan ancestry but tended to be more classical in the Libyan patients with higher frequency of myoclonic jerks and periodic EEG and a progressive course of shorter duration.

(Meiner et al. 1997) reviewed familial Creutzfeldt-Jakob disease with particular reference to the E200K mutation, which is unusually frequent in Libyan Jews.

The most common mutation associated with this disease is at codon 200 of PRN (Glutamic for Lysine, E to K) P. Interestingly, this mutation accounts for the three reported ethnogeographic clusters of CJD where the local incidence of CJD is around 100-fold higher than elsewhere (among Libyan Jews, in a region of Slovakia, and in a region of Chile) (Goldfarb et al. 1990a; Goldfarb et al. 1990b; Hsiao et al. 1991);(Brown et al. 1992).

Individuals homozygous for the 200 mutation have been identified and are phenotypically indistinguishable from heterozygotes, indicating that this condition is a fully dominant disorder (Hsiao et al. 1991). Patients with this condition have now been reported in several other countries outside the recognized clusters, including the UK.

1.2.3.2. Fatal familial insomnia (FFI)

Familial fatal insomnia is associated with a specific sequence of the prion protein gene (PRNP). FFI, an autosomal dominant disorder, is characterized by neuronal degeneration limited to selected thalamic nuclei and progressive insomnia. FFI is associated with the asp178-to-asn mutation of the PRNP gene (D178N) when the amino acid at position 129 is methionine; the same mutation in 178 results in Creutzfeldt-Jacob disease when the amino acid at position 129 is valine.

(Harder et al. 1999) presented a large German kindred with fatal familial insomnia. Molecular genetic analysis of the PRNP gene confirmed that the D178N mutation segregated with methionine at the polymorphic codon 129 in all 7 affected patients examined. The authors noted a wide spectrum of clinical presentations and emphasized the difficulty in establishing the diagnosis of fatal familial insomnia on clinical and pathologic grounds alone. They were unable to confirm the previously reported relationship between the status of the M/V polymorphism at codon 129 and age at onset of this disease

The only mutation associated to this form is at position 178 (on a Methionine allele at codon 129), that was first reported in 1986 (Lugaresi et al. 1986) associated with this disease.

1.2.3.3. Gerstmann-Straussler-Scheinker disease (GSS or GSD)

The disorder is caused by mutation in the prion protein gene (PRNP).

Gerstmann-Straussler-Scheinker disease is a rare familial disorder characterised by cerebellar ataxia, progressive dementia, and absent reflexes in the legs and pathologically by *amyloid* plaques throughout the central nervous system (Masters et al. 1981; Seitelberger 1981). Onset is usually in the fourth or fifth decade and in the early stages ataxia is predominant. Dementia develops later. The course ranges from 2 to 10 years. In addition to spinocerebellar and corticospinal tract degeneration, extensive *amyloid* plaques are found throughout the CNS, and in many cases, *spongiform degeneration* is found (Masters et al. 1981).

Like CJD and kuru, GSD is a form of subacute spongiform encephalopathy. Cases of GSD are clinically similar to the ataxic type of CJD. Although there are many neuropathologic similarities, GSS differs from CJD by the presence of kuru-plaques and numerous multicentric, floccular plaques in the cerebral and cerebellar cortex, basal ganglia, and white matter.

The most common mutation associated to this disease is at position 102, that was first reported in 1989 (Hsiao et al. 1989).

2. COMPARATIVE GENETICS BETWEEN HUMANS AND CHIMPANZEES

2.1. Introduction

Studies of genetic variation in close relatives of humans, such as chimpanzee, can help to identify which features of our evolution are peculiar to us and which are shared.

The draft chimpanzee sequence has been publicly available since November 2003 from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html). This second hominoid genome is the first complete nonhuman primate genome against which human genetic features can be compared. Specifically, a comparison of humans and other apes can help to identify targets of natural selection in the human lineage, compare demographic histories and assess the extent to which recombination and mutation rates are conserved among closely related species (Kaessmann et al. 1999b) (Huttley et al. 2000) (Nachman and Crowell 2000) (Kaessmann et al. 2001) (Yi et al. 2002) (Enard et al. 2002), (Wall 2003). In these respects, the study of the closest living relative of humans, the chimpanzee, may be especially enlightening (Olson and Varki 2003).

The estimation of divergence times between primate groups is more controversial than the phylogeny, largely due to the uncertainties associated with fossil calibration points. For example, the divergence of the lineage leading to lemurs and the lineage to all other primates has been dated to 63 mya

(Goodman et al. 1998) or to 80 mya (Springer et al. 2003). Using two different fossil calibration points and testing several statistical methods, Glazko & Nei (Glazko and Nei 2003) estimated the divergence times between humans and other primates. According to them, the human lineage diverged from that of the chimpanzee approximately 5-7 mya, that of the gorilla 6–8 mya, that of the orangutan 12–15 mya, that of Old World monkeys 21–25 mya, and that of New World monkeys 32–36 mya (Fig. 10). These dates accord with previous studies (Chen et al. 2001) (Goodman 1999), even if they are disputed by some (Arnason et al. 2000). The divergence times of DNA sequences from bonobos and chimpanzees are less well dated and range from 0.9 million years (Kaessmann et al. 1999b) to 2.5 million years (Gagneux et al. 1999), with the largest dataset collected so far suggesting 1.8 million years (Yu et al. 2003).

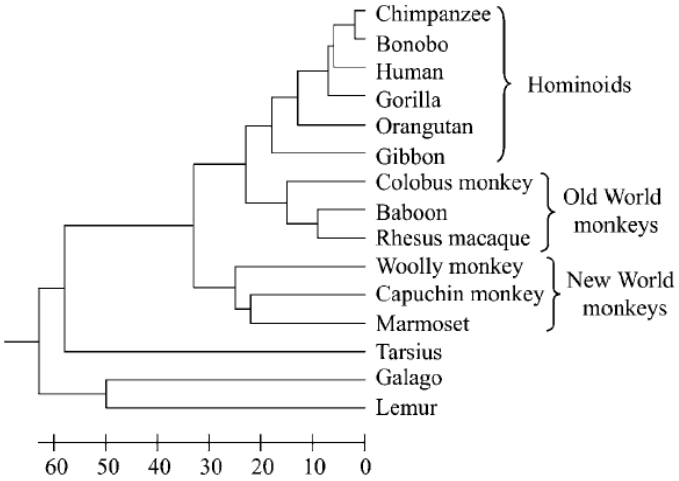


Figure 10. Phylogeny of a few primate species. Divergence times are indicated. From (Enard et al. 2002).

Chimpanzees are classified into two species: the common chimpanzee, *Pan*

troglodytes, and the pygmy chimpanzee or bonobo, *Pan paniscus*. Among common chimpanzees, three "subspecies" are recognized on the basis of their geographic distribution: *Pan troglodytes verus* in west Africa, *P. t. troglodytes* in central Africa and *P. t. schweinfurthii* in eastern Africa. Recently, a fourth subspecies, *P. t. vellerosus,* has been suggested in western Africa based on mitochondrial DNA divergence (Gonder et al. 1997). Very little is known about the distribution of variation among these subspecies or between species (Morin et al. 1994).

2.1.1. Differences in phenotype

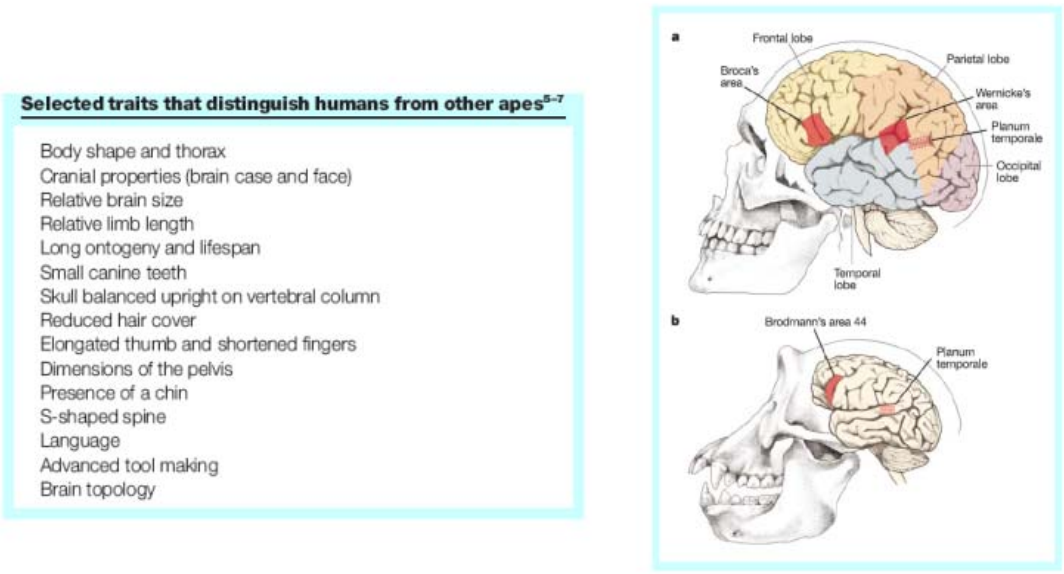


Figure 11. Traits that distinguish humans and other apes and Comparative neuroanatomy of humans and chimpanzees. . From (Carroll 2003)

From a rather extensive list of qualitative and quantitative features that distinguish humans from other apes (Fig. 11) (Carroll 2003), our large brain,

bipedalism, small canine teeth, language and advanced tool-making capabilities have been the focus of palaeoanthropology. The major physical traits are generally not singular elements, but entail concomitant changes in skeletal features involved in locomotion (for example, in the vertebral column, pelvis and feet, and in limb proportions), grasping (hand morphology and an opposable, elongated thumb) and chewing of food (the mandible and dentition), as well as life-history traits such as lifespan.

If we focus on comparative neuroanatomy of humans and chimpanzees what we can see is that both the human cranial capacity and the human brain are much larger. Two areas of the human brain that are associated with communication are shown (Fig. 11): Broca's area in the frontal lobe and Wernicke's area, which overlaps the posterior temporal lobe and parts of the parietal lobe. In the left hemisphere, Broca's area is larger, as is the planum temporale, which lies below the surface in Wernicke's area. These asymmetries have been found in corresponding regions of chimpanzee brains, suggesting that these areas in humans might be elaborations of a pre-existing communication centre in a common ancestor of apes and humans.

2.1.2 Biomedical significance

Humans and chimpanzees are similar with regard to many aspects of physiology and disease. Here, the focus is primarily on known and apparent differences (Table 3), because these are most relevant to the upcoming comparison of the human and chimpanzee genomes. The chimpanzee genome offers a unique opportunity to step back and look at the disease susceptibilities

of our species. Humans that live under modern conditions have a characteristic set of susceptibilities: numerous infectious diseases that are entirely human specific (or at least almost entirely); cardiovascular disease; carcinomas; obesity; type II diabetes; autoimmune diseases; major psychoses; and neurodegenerative diseases. Although individual humans vary in their genetic susceptibilities to these conditions, we should not allow these intra-species differences to distract attention from the larger question of why humans have this particular spectrum of susceptibilities. To some extent, the unique environments in which humans live undoubtedly contribute to human patterns of health and disease. However, the environments in which captive chimpanzees live increasingly resemble those of humans.

Medical condition	Humans	Great apes
Definite differences		
HIV progression to AIDS	Common	Very rare
Late complications in hepatitis B/C	Frequent	Uncommon
Plasmodium falciparum malaria	Susceptible	Resistant
Menopause	Universal	Rare
Likely differences		
Influenza A symptomatology	Moderate to severe	Mild
Alzheimer disease pathology (neurofibrillary tangles)	Common	Rare
Myocardial infarction	Common	Uncommon
Simian foamy virus infection	Very rare	Very common
Epithelial cancers (carcinomas)	Common	Rare
Escherichia coli K99 gastroenteritis	Resistant	Sensitive?

Table 3. Apparent biomedical differences between humans and great apes.

From (Olson and Varki 2003)

Hence, genetic differences are likely to contribute to differences in the typical health profiles of humans and captive chimpanzees, and the identification of the critical differences would be of great biomedical interest. Our existing

knowledge of the disease profiles of chimpanzees is fragmentary but intriguing. Table 3 summarizes examples of chimpanzee–human differences that are either well established or indicated by existing reports (Varki 2000). In addition to the conditions listed in (Table 3), anecdotal reports indicate that some common human conditions are rare in chimpanzees. These include early fetal wastage, hydatiform molar pregnancy, bronchial asthma, acne vulgaris, major psychoses, and autoimmune diseases such as systemic lupus, rheumatoid arthritis and multiple sclerosis (Varki 2000).

A striking feature of this list is that it includes many of the common diseases of central importance to contemporary biomedical research in economically developed countries. The medical goal would be to learn how to emulate aspects of primate biology that have been lost in humans. This concept has the potential to open a new front in biomedical research.

2.2 Molecular differences

2.2.1 Dynamic rearrangements

Major cytogenetic changes between humans and chimpanzees include three in the hominid lineage and seven in the panid lineage (Yunis and Prakash 1982). In addition, there are large numbers of minor changes that appear to be mediated by mobile elements and segmental duplications. Recently, it has been suggested that chromosomal rearrangements have a potential role to play in driving sequence divergence and speciation, based on the fact that

chromosomal regions with rearrangements appear to show higher divergences (Navarro and Barton 2003).

A study (Fortna et al. 2004) of genome-wide gene duplications in hominoids based on cDNA samples reported a bias for increase in copy number among genes with a lineage-specific difference in copy number in the human lineage (143 out of 140 differences).

2.2.2 Differences in tandem repeats regions

It has been suggested that contraction and expansion of tandem repeats affect human and chimpanzee genomes differentially (Cooper et al. 1998). It appears that triplet repeat expansion (e.g. polyglutamine tracts) in genes that are associated with several human diseases, such as spinocerebral ataxia, show variation in repeat length in chimpanzees versus humans, predisposing humans to the pathogenic effects of extreme expansion (Andres et al. 2003).

2.2.3 Inactivation or gain of genes in the human lineage

At the genetic level, an increasing number of genes that are inactivated in humans but still intact in the great apes are being identified. These include the genes encoding the T-cell receptor (TCRGV10 (Zhang et al. 1996), the sialic acid-modifying enzyme cytidine monophospho-N-acetyl-neuramic acid hydroxylase (CMAH (Chou et al. 1998), the endogenous sialic acid-binding immunoglobulin-like lectin 1 (SIGLECL1 (Goodman and Check 2002), keratin hair acidic pseudogene 1 (KRTHAP1 (Winter et al. 2001), myosin heavy

polypeptide 16 (MYH16 (Stedman et al. 2004), as well as numerous olfactory genes (Gilad et al. 2003) (Fig. 12).

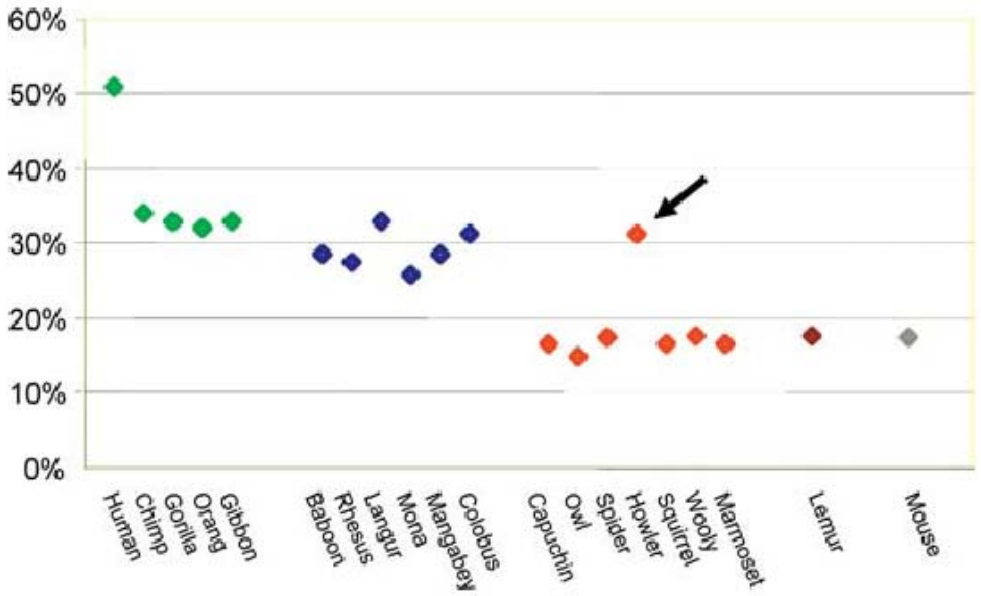


Figure 12. The proportion of OR pseudogenes in 20 species. Primate species are color-coded according to family. From (Gilad et al. 2003)

The sample size is small but the analysis of the chimpanzee genome will shed light on whether increased gene loss is characteristic of the hominid lineage. By contrast, there is only one reported gain of a functional gene in humans, the Y-linked protocadherin 11 gene (PCDH11Y (Crow 2002). Recent loss of gene function in humans might be due to mutations, such as Alu element-mediated deletions (e.g. in CMAH (Hayakawa et al. 2001).

2.2.4 Nonsynonymous/synonymous substitution ratios (Ka/Ks)

The previous examples given have a low effect on the divergence of these two genomes (human and chimpanzee). It is more likely that small differences in several genes account for the specific traits we observe between these two

species. Looking at specific changes in the gene between species can give a lot of information, as does nonsynonymous:synonymous substitution ratio (Ka:Ks). This Ka:Ks ratio has been used as a common measure of selection.

The number of reported instances of adaptive evolution based on Ka:Ks ratio calculation for whole genes is rapidly increasing. Although providing interesting candidates for genes involved in lineage-specific adaptations, these cases will have to be studied for their effects on function.

In a recent study (Dorus et al. 2004) , the molecular evolution of an extensive set of nervous system-related genes in primates was examined. They demonstrated that the average rate of protein evolution as scaled to neutral divergence (i.e., the *Ka*/*Ks* ratio) is significantly higher in primates than in rodents. One possible interpretation is the adaptive evolution of these genes in primates, but it could also be due to relaxed functional constraint. Nervous system genes with developmentally biased functions displayed much greater primate-rodent *Ka*/*Ks* disparity than the entire set of genes. In contrast, the *Ka*/*Ks* of genes that function predominantly in the routine physiological operations and maintenance of the nervous system showed much less primate-rodent disparity. The latter observation argues against reduced functional constraint on the primate nervous system per se, and together, these results are more consistent with the notion of adaptive evolution. Mutations in many nervous system genes, including those with significantly higher *Ka*/*Ks* in primates, have been shown to cause severe nervous system defects in humans. The above results argue against the possibility of relaxed functional constraint on the primate nervous system. Instead, they are more consistent with the

interpretation that higher *Ka*/*Ks* of nervous system genes in primates—especially along the lineage leading to humans—is a reflection of adaptive evolution.

2.2.5 From genomic architecture to gene expression

A more difficult level of investigation is the level of gene expression, where shades of gray (subtle changes in expression) rather than black-white (function or dysfunction of gene) might be determining species-specific traits.

Comparative gene expression and proteomic studies in tissues of humans and great apes (Uddin et al. 2004) (Gagneux et al. 2001) (Enard et al. 2002) (Caceres et al. 2003) (Figs. 13 and 14) (Karaman et al. 2003) have provided evidence in humans for higher rates of change in expression, and for the general up regulation of numerous genes in brain tissue, when compared with non-central nervous system tissues. These expression studies have helped identifying genes that are linked to functional or metabolic categories, and will trigger studies of the candidate genes and the metabolic pathways that they affect.

Preliminary data suggest that gene expression patterns of the human brain might have evolved rapidly (Enard et al. 2002) (Caceres et al. 2003) (Uddin et al. 2004).

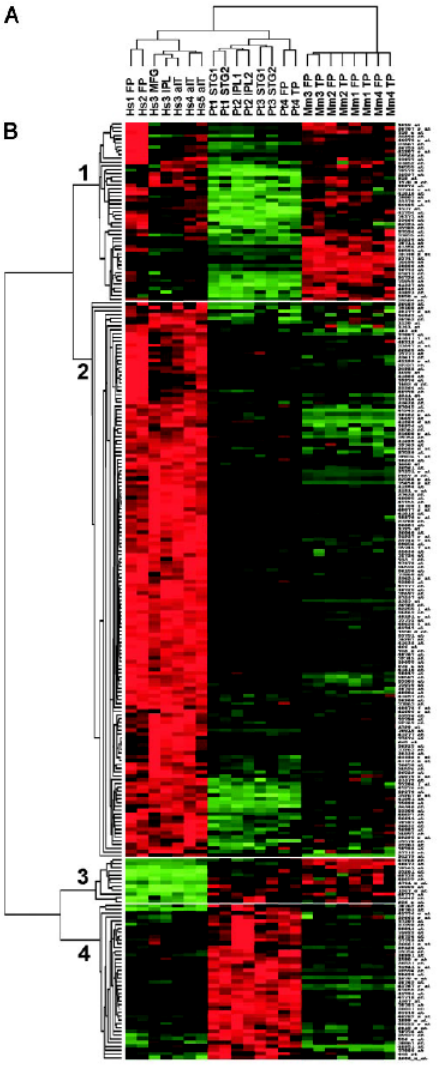


Fig. 13 (left). Gene expression analysis of human, chimpanzee, and rhesus macaque cerebral cortex. (*A*) Dendrogram showing the hierarchical clustering of the different cortex samples according to the hybridization signals of 9,733 probe sets detected in at least one of the samples from any species (Hs1–5, humans; Pt1–4, chimpanzees; Mm1–4, rhesus macaques). (*B*) Hybridization levels for 246 probe sets that show differences in signal intensity between human and chimpanzee cortex. From (Caceres et al. 2003).

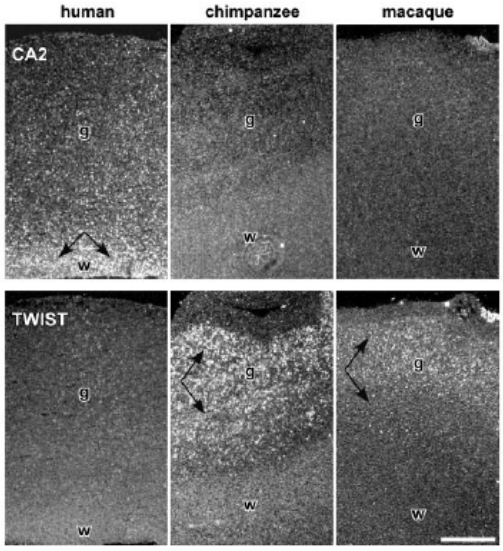


Fig. 14 (right). Histological study of gene-expression differences in the cortex between humans and non-human primates. *In situ* hybridization confirms that CA2 is expressed at higher levels in the cortex of humans than in chimpanzees or rhesus macaques (*Upper*). TWIST is weakly expressed in human cortex compared with chimpanzees and macaques (*Lower*). From (Caceres et al. 2003).

2.2.6 Recombination rates between humans and chimpanzees

An increasing number of studies are being carried out of variation in recombination rate along the genome, but little is known about how these rates change temporally. In humans, recombination rates differ among individuals (Kong et al. 2002) (Cullen et al. 2002), and this variation, at least in females,

seems to be heritable (Kong et al. 2004). These observations suggest that recombination rates could evolve over time. At the fine scale, two recent studies suggest that human hotspots are not always conserved in other primates: the *B*-globin hotspot seems to be absent in rhesus macaques (*Macaca mulatta*) (Wall et al. 2003), and the TAP2 hotspot seems to be absent in chimpanzees (*Pan troglodytes*) (Ptak et al. 2004). These results suggest that fine-scale recombination can evolve rapidly, it is difficult to generalize because they are based on only two hotspots and the two regions were known *a priori* to contain a human hotspot. To find a general view of the recombination rate comparing the two species a study of two regions totalling 14 Mb of genomic sequence in chimpanzees and the orthologous regions in humans was realized (Ptak et al. 2005). A statistical approach described in a previous study (Crawford et al. 2004) was used to infer recombination rates. The main conclusions of this work by (Ptak et al. 2005) (Fig. 15) were:

-Recombination hotspots are not conserved among the two species.

-They rejected the hypothesis that total recombination rates at the 50-kb scale are correlated in humans and chimpanzees ($P<0.01$)

-When examining the correlation in the background recombination rate for the 50-kb windows, they found that the estimated background rates of recombination were also significantly but weakly correlated between the two species.

-Therefore, because background rates are weakly conserved but hotspots are not, they hypothesize that the conservation in total recombination rate reflects the conservation in background recombination rates.

-lack of concordance between hotspots raises the question of how hotspots change so quickly between species with such high sequence similarity.

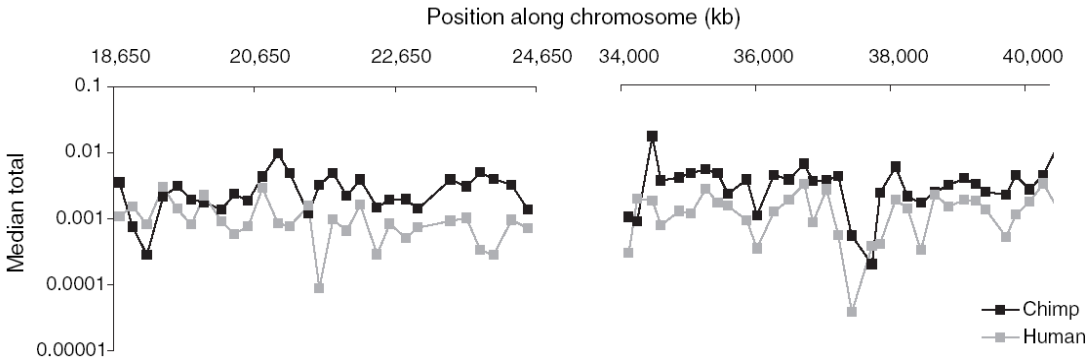


Figure 15. Concordance of total p for chimpanzee and human samples. Total recombination rates, as estimated at the 50 kb scale, are significantly but weakly correlated in humans and chimpanzees. From (Ptak et al. 2005).

Similar results were obtained in another study (Winckler et al. 2005). Recombination hotspots were found rarely (if at all) at the same positions in the two species, and no correlation was observed in estimates of fine-scale recombination rates.

2.2.7. Differences in the intraespecific diversity

The next obvious question is whether humans or chimpanzees are exceptional among primates in having low and high amounts of DNA sequence diversity, respectively.

The level of variability in a population or species is influenced by the demographic history of the group and also, it is influenced by the locus studied. One of the most difficult aspects in the study of chimpanzees is ensuring a big enough and representative sample to study the variability of chimpanzee

species. As this species is in danger, and the countries where this species is found are not easily accessible, this turns to be very complicated.

One study that reached these objectives was based on mitochondrial DNA sequences and published by (Gagneux et al. 1999)(Fig. 16), where they observed that the level of diversity was greater in this species than in humans. This was in accordance with a previous study by (Ferris et al. 1981).

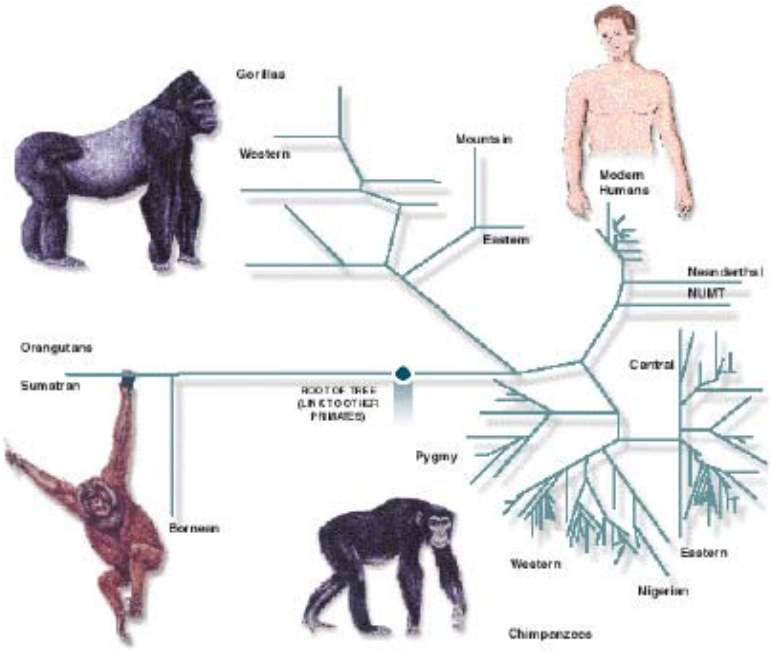


Figure 16. Unrooted phylogram of the neighbor-joining tree of 1.158 different control region sequences (mithocondrial) (Gagneux et al. 1999).

Another non recombinant segment in the genome is the NRY region of chromosome Y. Stone et al 2002 also indicated that humans are the species showing least variation when compared with other primates and with a more recent common ancestor (MRCA).

This conclusion agrees with other nuclear loci, (Kaessmann et al. 1999a; Kaessmann et al. 1999b; Yu et al. 2003) suggesting that the population history of humans is different at least from chimpanzees, bonobos and gorillas (Kaessmann et al. 2001).

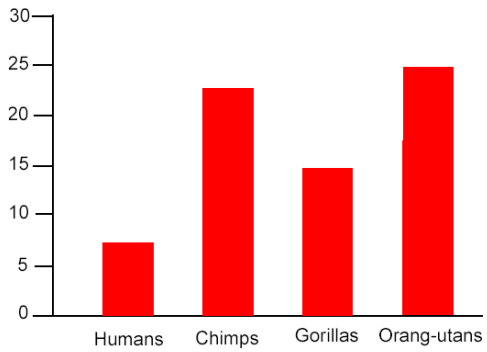


Fig. 17. DNA sequence diversity within humans and great apes. Values are based on the number of variable positions within each species taking the number of sequences determined into account. From (Kaessmann et al. 1999a).

However the first studies that analyze the variability in each species used STR (short tandem repeats) nuclear loci, and these particular elements show bigger diversity in humans than in chimpanzee (Wise et al. 1997). But the rapid evolution of these elements are particular, different groups of STRs showing differences in variability. (Garza et al. 1995) and (Crouau-Roy et al. 1996) indicated that maybe there is not a clear and general directionality of these repetitive loci.

2.2.8 Differences in effective population size (Ne)

Humans have less genetic variation than do their closest relatives. This is, at first glance, perhaps surprising. Based on population genetic theory, levels of

genetic variation within species should correlate positively with population size. This predicted correlation comes about because the strength of random genetic drift - which results in the loss of genetic variation - increases at lower population sizes. Yet, the human population numbers in the billions and the population sizes of chimpanzees and gorillas is fewer than a hundred thousand in each case.

What could explain this discrepancy? The strength of genetic drift is dependent not on the current census population size but on the historical population sizes. The relatively low levels of genetic variation in humans can be explained by a severe population bottleneck, where the population of our species was likely reduced to a few thousand. It could also be explained by a more moderate, sustained bottleneck. During this bottleneck, the population was possibly in the tens to hundreds of thousands for a more considerable time. Alternately, natural selection could also either increase or decrease the extent of variation in one of the species. Yet, because it is unlikely that natural selection would act in the same way on multiple regions of the genome, the difference in the extent of genetic variation between humans and chimpanzees is more likely a consequence of historical demography.

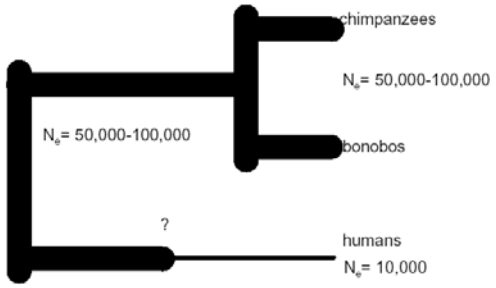


Figure 18. Effective population size for humans and chimpanzees.

3 Signatures of natural selection in the human genome

3.1. Introduction

Natural selection is defined as the differential reproduction of genetically distinct individuals or genotypes within a population. Differential reproduction is caused by differences among individuals in such traits as mortality, fertility, fecundity, mating success, and the viability of the offspring. When a population consists of individuals that do not differ from one another in such traits, it is not subject to natural selection. Selection may lead to changes in allele frequencies over time. However, a mere change in allele frequencies from generation to generation does not necessarily indicate that selection is at work. Other processes, such as random genetic drift, can bring about temporal changes in allele frequencies as well. The opposite is also true: a lack of change in allele frequencies does not necessarily indicate that selection is absent.

Most new mutations arising in a population reduce the fitness of their carriers. Fitness is defined as a measure of the individual´s ability to survive and reproduce. Such mutations are called deleterious, and they will be selected against and eventually removed from the population. This type of selection is called negative or purifying selection. Occasionally, a new mutation may be as fit as the best allele in the population. Such a mutation is selectively neutral, and its fate is not determined by selection. In rare cases, a mutation may arise

that increases the fitness of its carriers. Such a mutation is called advantageous, and it will be subjected to positive or advantageous selection.

3.1.1 Hardy-Weinberg law

Suppose that at a locus, A1 is the common allele (the wild type), and A2 is the mutant type (say, a disease-causing mutation). It is conventional to use p and $q = 1 - p$ to represent the allele frequencies in the population. The Hardy-Weinberg law states that if a diploid randomly-mating population is in equilibrium, and there is no selection, mutation or migration, the genotype frequencies are specified by the allelic frequencies according to the binomial probabilities:

Genotype	A1A1	A1A2	A2A2
Frequency	p^2	$2pq$	q^2

In real data, the observed genotype frequencies might departure from the prediction from the H-W law, and many factors might be responsible for the departure, such as population subdivision, non-random mating, or natural selection.

3.1.2 Selection

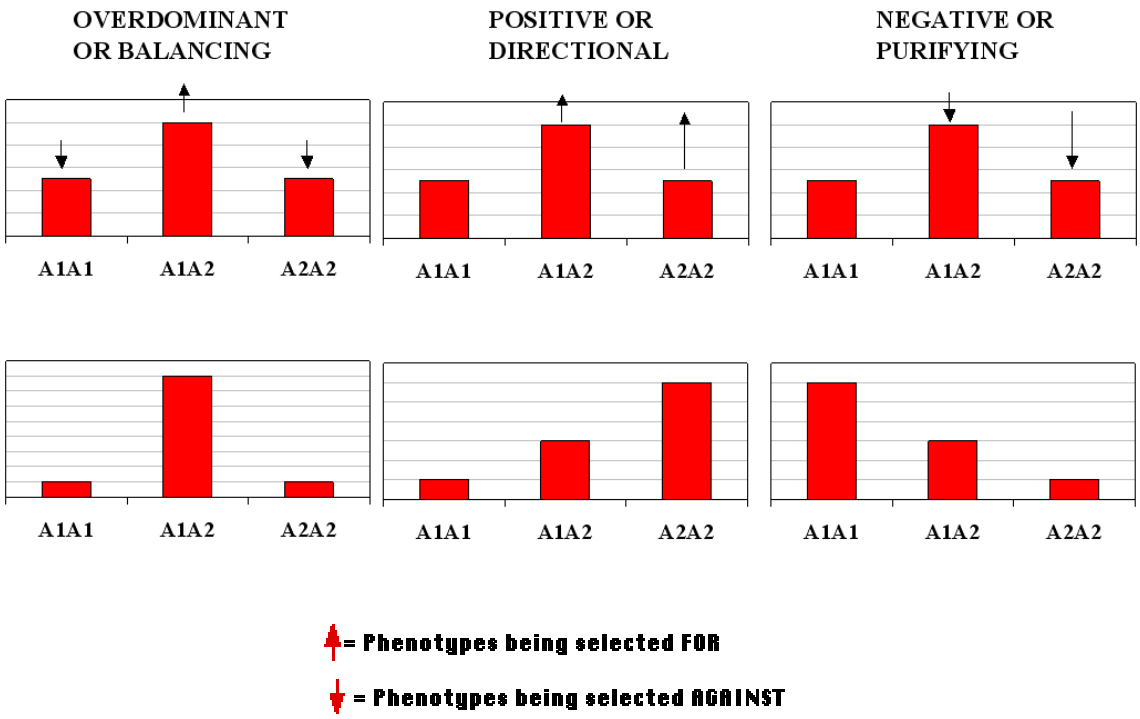
The theory of evolution by natural selection is based on the following premises or observations.

(1) There is reproductive excess; that is, the population size would increase without limit if all individuals born survived. Since no population can increase boundlessly, not all individuals will survive;

(2) There is difference among individuals in their chances of survival and reproduction;

(3) Such differences in *fitness* are due to the genetic composition of individuals.

Evolution will then be inevitable. Natural selection means that individuals carrying certain alleles are better adapted to the environment and have better chances for survival and reproduction, and thus contribute more to the next generation. In population genetics, we use the selection coefficient (*s*) of an allele to measure its fitness relative to the wildtype allele.



(a) ***Positive or adaptive selection*** *(s > 0, Darwinian selection, adaptive evolution)* favouring a dominant allele.

Genotype	A1A1	A1A2	A2A2
Relative fitness	*1*	1 + *s*	1 + 2*s*

The mutant allele A2 has a selective advantage and is under *positive* selection. A2 will spread into the population and reach fixation while A1 will be lost.

(b) ***Purifying or negative selection*** *(s < 0, negative selection, selective constraint)* against a dominant allele.

Genotype	A1A1	A1A2	A2A2
Relative fitness	*1*	1 - *s*	1 - 2*s*

In this model, A2 is deleterious and is under *negative* or *purifying* selection. The new allele A2 will be purged by natural selection from the population. Most proteins appear to be near optimal, perfected after millions of years of evolution. As a result, most new mutations tend to disrupt the structure and function of the protein and are deleterious. The protein is said to be under *selective constraint*.

(c) ***Overdominance or balancing selection:*** The heterozygote A1A2 has a higher fitness than either homozygote (A1A1 or A2A2). The system has a stable equilibrium at $p = s2/(s1 + s2)$.

Genotype	A1A1	A1A2	A2A2
Relative fitness	$1 - s_1$	1	$1 - s_2$

The classic case is the sickle-cell anemia in humans, which is prevalent in populations affected by malaria caused by the mosquito-borne parasite *Plasmodium falciparum*.

3.1.3 The neutral and nearly neutral theory

The consensus reached from the theoretical work of R.A. Fisher, J.B.S. Haldane, and S. Wright in the 1930s and generally accepted in the 1950s or 1960s was that natural selection was the driving force of evolution, and random genetic drift was unimportant. Most populations were believed to be quite homogeneous with little variation. Most new mutations were deleterious and were quickly removed from the population. Occasionally an advantageous mutation occurred and spread over the entire population. However, starting from 1966, surprisingly high levels of genetic variation were detected in allozymes using electrophoresis. The *neutral theory*, or more exactly, *the neutral-mutation random-drift hypothesis*, was proposed in 1968 and 1969 by Kimura mainly to accommodate this surprising funding.

The neutral theory asserts that the genetic variation we observe today, both the polymorphism within a species and the diversity between species, is due to random fixation of neutral mutations rather than fixation of advantageous mutations by natural selection. When the selective coefficient $|s| << 1/N$, or

when $|Ns| << 1$, the fate of the allele is largely determined by genetic drift. Note that the neutral theory maintains that most of new mutations are deleterious and are removed from the population by *purifying selection*.

These mutations do not contribute to the genetic variation we see today. A very small proportion of new mutations are advantageous. The rest of the mutations have fitness effect so small that natural selection has no role in determining their fate. These neutral mutations are responsible for most of the observed variation according to the theory.

The strictly neutral theory accounts for two classes of mutations: the lethal or highly deleterious mutations, which do not contribute to the genetic variation at all, and the completely neutral mutations for which selection has no effect on their fate.

The neutral theory claims or predicts that

(1) Most mutations are deleterious and are removed by *purifying selection*;

(2) The nucleotide *substitution rate* is equal to the neutral mutation rate (total mutation rate times the proportion of neutral mutations); If the neutral mutation rate is constant among species (either in calendar time or in generation time), the substitution rate will be constant. This prediction provides an explanation for the molecular clock hypothesis;

(3) Functionally more-important genes or gene regions change more slowly.

(4) Within-species polymorphism and between-species divergence are two phases of the same process.

The neutral theory does not claim that all mutations are neutral.

The past three decades of population genetics have focused on testing the neutral mutation hypothesis and on extending it or developing alternative models. The controversy surrounding the neutral theory has generated a rich body of population genetics theory and analytical tools. The nice thing about the hypothesis is that it makes explicit predictions about the data, so that it is possible to test the theory using real data.

The strict neutral model of Kimura has often been rejected. Tomoko Ohta (1973) proposed the *slightly deleterious mutation hypothesis*, which allows for mutations whose effects are slightly deleterious. Natural selection will have some role in determining the fate of such mutations, and the fixation probability is positive but smaller than for neutral mutations.

Note that the model does not allow for slightly advantageous mutations. The scene of the world is bleak, with essentially all mutations being harmful. Later this theory changed to the *nearly neutral hypothesis*, which allows for both slightly deleterious and slightly advantageous mutations. In contrast to the strict neutral model, in which the dynamics depends on the neutral mutation rate alone and not on other factors such population size and selective pressure, the slightly deleterious mutation or nearly neutral theories involve all these parameters and are very difficult to refute. This is also the case for various selection models (Gillespie 1989).

3.1.4 Genealogy and Selection

Different gene genealogies can be seen (Bamshad and Wooding 2003) depending on the type of selection that is acting : (Fig.19):

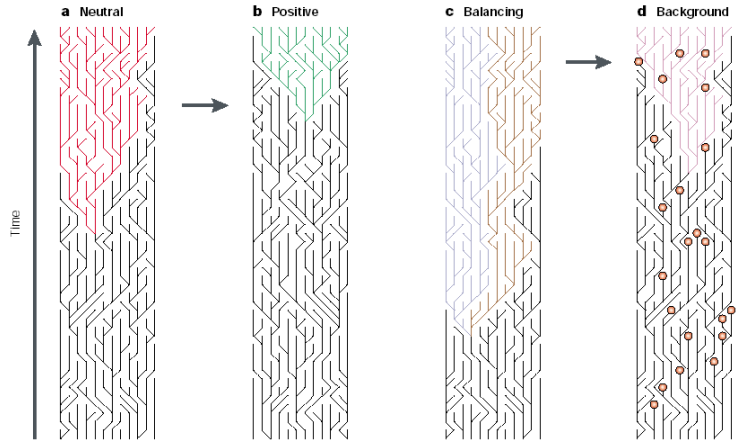


Figure 19. Effects of natural selection on gene genealogies and allele frequencies Each panel (**a**–**d**) represents the complete genealogy for a population of 12 haploid individuals From (Bamshad and Wooding 2003).

a | The genealogy of a NEUTRAL (no selection) allele (red) as it drifts to fixation.

b | The genealogy of an allele (green) that is driven to fixation more quickly after the onset of POSITIVE SELECTION (arrow) compared with expectations under a neutral model. Note that the genealogy has a more recent coalescence.

c | The genealogy of two alleles (blue and gold) under BALANCING SELECTION, which are driven neither to fixation nor to extinction. As a result, the genealogy of the two alleles has an older coalescence.

d | The genealogy of an allele (purple) that drifts to fixation under the influence of BACKGROUND OR PURIFYING SELECTION. Each circle represents the elimination of a deleterious mutation by background selection. The coalescence of the lineage is more recent than expected under a neutral model because a linked deleterious mutation caused the extinction of one lineage (arrow) more quickly than would be expected.

3.2 Statistical methods used to detect selection

A low level of genetic diversity within a population may reflect limited immigration, extensive drift, or selective pressures against a certain set of alleles. Likewise, a high level of genetic diversity may result from extensive immigration, a large population size, or selection favouring the increase of genetic diversity. In each case these factors can be combined together. Therefore, before trying to explain diversity at a locus in terms of ancient population processes, we need to account for any effect of selection. As seen before there are different types of selection that account for different effects on genetic diversity. Another important point to take into account is that selection can act specifically on a determined locus, but also, selection could be operating on linked loci.

We can test for the footprint of past selection by different methods (Fig. 20), which typically compare some feature of the observed diversity to that expected under neutral evolution. Such methods are also known as neutrality tests. The perfect test to check this hypothesis of neutrality does not exist, as different methods require different sorts of data and vary in their ability to detect the influence of different selective regimes. Most of these methods are oriented towards summary statistics of DNA sequence diversity.

An important remark is that a significant difference between any test statistic and neutral expectations need not result solely from selection. Demography

needs to be taken also in consideration. The equilibrium model of neutral evolution also assumes that the population has certain demographic characteristics, and one of the assumptions in the neutral model is that our species has been in constantly sized neutral equilibrium, which it has not.

In the following sections I am going to classify the tests on the basis of the different effects of selection each attempts to detect. The power of these tests to detect selection depends on, among other factors, the characteristics of the selective regimes: such as the type of selection operating; the strength of selection; the period during which selection occurred or is occurring.

Test	Compares
Tests based on allelic distribution and/or level of variability	
Tajima's *D*	The number of nucleotide polymorphisms with the mean pairwise difference between sequences
Fu and Li's *D*, *D**	The number of derived nucleotide variants observed only once in a sample with the total number of derived nucleotide variants
Fu and Li's *F*, *F**	The number of derived nucleotide variants observed only once in a sample with the mean pairwise difference between sequences
Fay and Wu's *H*	The number of derived nucleotide variants at low and high frequencies with the number of variants at intermediate frequencies
Tests based on comparisons of divergence and/or variability between different classes of mutation	
d_N/d_S , K_a/K_s	The ratios of non-synonymous and synonmyous nucleotide substitutions in protein coding regions
HKA	The degree of polymorphism within and between species at two or more loci
MK	The ratios of synonymous and non-synonymous nucleotide substitutions in and between species

HKA, Hudson–Kreitman–Aguade; MK, McDonald–Kreitman.

Figure 20. Tests commonly used to detect selection. From (Bamshad and Wooding 2003)

3.2.1 Tests based on allelic distribution or levels of variability

3.2.1.1. One locus

3.2.1.1.1. Tajima´s D

One of the milestones of population genetical theory was the proposal of the Ewens sampling formula (Ewens 1972) (Fig. 21). This formula provides an analytical expression for different allele frequencies (usually called allelic partition) under the infinite allele model, whereby every mutation is to a new allelic type, for a sample obtained from a single population of constant size with no population structure.

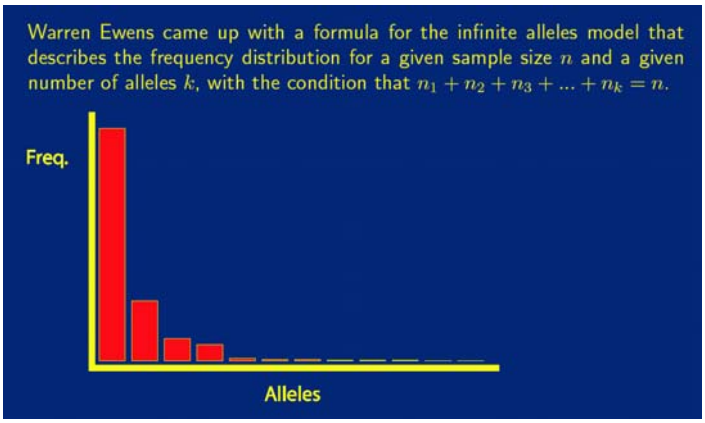


Figure 21. Ewens sampling formula.

Using Ewens's sampling formula, one of the most famous tests of neutrality, the Ewens-Watterson test (Ewens 1972; Watterson 1977) (Fig. 22) was developed. In this test the expected homozygosity, given the observed number of alleles, is compared to the observed homozygosity.

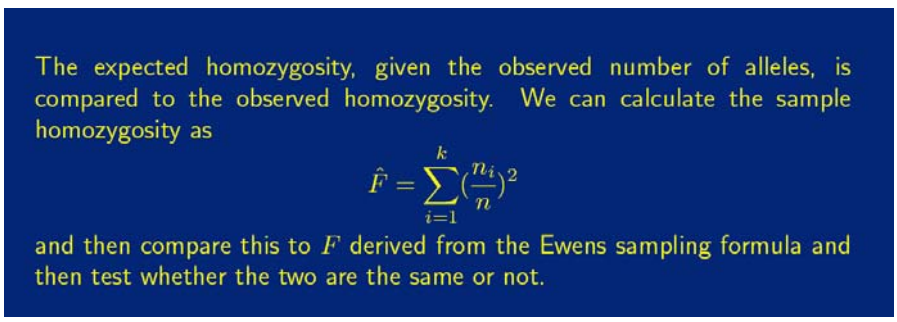


Figure 22. Ewens-Watterson test for neutrality

If the difference between the observed and expected homozygosity is larger than some critical value, the neutral null hypothesis can be rejected. The problems with this test is that it assumes that the allele distribution is at equilibrium and the other problem is that all alleles need to be identified. This test is applicable to data for which the infinite-alleles model might be reasonable.

For nucleotide data, one of the most popular tests is Tajima's D-test (Tajima 1989). Tajima's D is the scaled difference in the estimate of $\theta = 4N_e\mu$ (N_e = effective population size, μ = mutation rate per generation) based on the number of pairwise differences and the number of segregating sites in a sample of nucleotide sequences. It is defined as

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_\omega}{S_{\hat{\theta}_\pi - \hat{\theta}_\omega}}$$

where $\hat{\theta}_\pi$ is an estimator of θ based on the average number of pairwise differences, $\hat{\theta}_\omega$ is an estimator of θ based on the number of segregating sites and $S_{\hat{\theta}_\pi - \hat{\theta}_\omega}$ is an estimate of the standard error of the difference of the two estimates. If the value of D is too large or too small the neutral null hypothesis is rejected. The critical values are obtained by simulations if mutation rate variation and recombination are taken into account.

Segregating sites within a set of sequences are present at different frequencies; some are found in only a single sequence (singletons), while others are in multiple sequences. Both selection and demography shape the spectrum of

these frequencies (known as the site frequency spectrum). For example, in a population that has undergone a recent expansion many variant sites will have arisen relatively recently and so will be present only at low frequencies at all loci. Such a population can be considered to have an excess of rare alleles. Similary, if a specific lineage at a locus is undergoing positive selection, it will be increasing in frequency relative to other lineages, and so a lineage-specific excess of rare alleles may be observed at that locus, but not at others. In contrast, population subdivision and balancing (overdominant) selection maintain multiple lineages in the population for longer than would be expected under neutral evolution, and so produce an excess of intermediate frequency alleles. Population subdivision increases the number of intermediate frequency alleles at all loci, whereas balancing selection only influences the site frequency spectrum at the locus at which selection is acting.

Methods addressing the frequencies of variant sites are often based on the expectation that under neutral evolution different estimates of θ should be equal. Only some estimates of θ incorporate information on allele frequencies, the number of segregating sites (S) is independent of frequencies, but nucleotide diversity (π) is not. Consequently discrepancies between estimates of θ that incorporate frequency information differently (or not at all) detect departures from neutral expectations of the allele frequency spectrum. Tajima´s D compares two estimates of θ, based respectively, on S and π. Under neutrality, Tajima´s D is expected to be zero. Significantly positive values

of this statistic indicate population subdivision or balancing selection, whereas negative values indicate positive selection, purifying selection or population growth (Table 4).

TAJIMA´S D			
	SITE FREQUENCY SPECTRUM	**DEMOGRAPHY**	**SELECTION**
SIGNIFICANTLY POSITIVE VALUES $\hat{\theta}_\pi >> \hat{\theta}_\omega$	Excess of intermediate polymporphisms	Population subdivision	Balancing Selection
SIGNIFICANTLY NEGATIVE VALUES $\hat{\theta}_\pi << \hat{\theta}_\omega$	Excess of rare polymorphisms	Population growth	Positive Selection OR Purifying Selection

Table 4. Tajima´s D values interpretation.

As we will see Tajima´s D test has also some problems:

- Null-hypothesis includes several assumptions such as constant population size and no population structure
- It is difficult to distinguish between selection and demography.
- It is difficult to distinguish between positive and purifying selection (both with negative D values)

Other one locus tests based on allelic distribution or levels of variability

There are several similar tests based on slightly different test statistics such as the tests by (Fu and Li 1993), (Simonsen et al. 1995) and (Fay and Wu 2000). A likelihood ratio test of a similar problem was described in (Galtier et al. 2000).

3.2.1.1.2 Fu and Li

A new approach of neutrality test was proposed by (Fu and Li 1993). Considering the distribution of the mutations in the genealogy of a random sample of genes from the population. "Old" mutations will tend to be found in the older part of the genealogy while "new" mutations will likely be found in the younger part of the genealogy. The older part of the genealogy consists mainly of internal branches, while the younger part mainly of external branches. In the presence of purifying or negative selection there will tend to be an excess of mutations in the external branches because deleterious alleles are present in low frequencies. Also there is likely to be an excess of mutations in the external branches if an advantageous allele was recently fixed in the population, because then the majority of the mutations in the population are expected to be young. On the other hand, if balancing (overdominant) selection is operating at the locus, then some alleles may be old and so there may be deficiency of mutations in the external branches. Therefore, comparing the numbers of mutations in internal and external branches with their expectations under selective neutrality should be a powerful way to detect selection. This is the idea behind the proposed tests in their approach. The tests are known as Fu

and Li with an outgroup (Fu and Li´s F and Fu and Li´s D). Where the outgroup is not used, then, they received the name of Fu and Li´s F* and Fu and Li´s D*.

D* test statistic

The D* test statistic is based on the differences between ηs, the number of singletons (mutations appearing only once among the sequences), and η, the total number of mutations ((Fu and Li 1993), p. 700 bottom).

$$D^* = \frac{\left(\frac{n}{n-1}\right)\eta - a_n\eta_s}{\sqrt{u_{D^*}\eta + v_{D^*}\eta^2}}$$

F* test statistic

The F* test statistic is based on the differences between ηs, the number of singletons (mutations appearing only once among the sequences), and k, the average number of nucleotide differences between pairs of sequences ((Fu and Li 1993), p. 702; see also Simonsen et al. 1995, equation 10).

$$F^* = \frac{\Pi_n - \frac{n-1}{n}\eta_s}{\sqrt{u_{F^*}\eta + v_{F^*}\eta^2}},$$

D test statistic

The D test statistic is based on the differences between ηe, the total number of mutations in external branches of the genealogy, and η, the total number of mutations ((Fu and Li 1993) , equation 32).

$$D = \frac{\eta - a_n \eta_e}{\sqrt{u_D \eta + v_D \eta^2}},$$

F test statistic

The F test statistic is based on the differences between ηe, the total number of mutations in external branches of the genealogy, and k, the average number of nucleotide differences between pairs of sequences ((Fu and Li 1993), p. 702, top).

$$F = \frac{\Pi_n - \eta_e}{\sqrt{u_F \eta + v_F \eta^2}},$$

The D and F test statistics can also be computed using S, the number of segregating sites instead of η, the total number of mutations (see Simonsen et al. 1995). Under the infinite sites model (with two different nucleotides per site) both D and F values should be the same (S and η have the same value). However, if there are sites segregating for more than two nucleotides, values of S will be lower than those of η.

The major difference between Tajima´s D test and Fu and Li´s tests is that Tajima uses the difference between η and Π_n, whereas these tests

use whether the difference between ηi and ηe or the difference between ηe and $\prod_n$.

3.2.1.1.3 Fay and Wu´s test

Positive selection can be inferred from its effect on linked neutral variation. In the restrictive case when there is no recombination, all linked variation is removed. If recombination is present but rare, both deterministic and stochastic models of positive selection show that linked variation hitchhikes to either low or high frequencies. While the frequency distribution of variation can be influenced by a number of evolutionary processes, an excess of derived variants at high frequency is a unique pattern produced by hitchhiking (derived refers to the nonancestral state as determined from an outgroup). (Fay and Wu 2000) adopted a statistic, *H*, to measure an excess of high compared to intermediate frequency variants. Only a few high frequency variants are needed to detect hitchhiking since not many are expected under neutrality. This was of particular utility in regions of low recombination where there is not much variation and in regions of normal or high recombination, where the hitchhiking effect can be limited to a small region.

The H statistic (Fay and Wu 2000) is based on the differences between two estimators of θ: θ_π (or k), the average number of nucleotide differences between pairs of sequences, and θ_H, an estimator based on the frequency of the derived variants (as opposed to the ancestral variants):

$$\hat{\theta}_{\pi} = \sum_{i=1}^{n-1} \frac{2S_i i(n-i)}{n(n-1)}$$

$$\hat{\theta}_{W} = \left(\sum_{i=1}^{n-1} \frac{1}{i}\right)^{-1} \sum_{i=1}^{n-1} S_i$$

$$\hat{\theta}_{H} = \sum_{i=1}^{n-1} \frac{2S_i i^2}{n(n-1)}$$

An outgroup was used to infer the derived and ancestral states for all polymorphism data analyzed. However, a backmutation would result in the incorrect inference of the derived polymorphic state. As compensation, the probability of misinference was incorporated into the null distribution of the *H* statistic by exchanging the frequency of the derived and ancestral state, with probability equal to that of misinference for each segregating site.

Selection versus demography

These tests have had great success in many applications in testing the neutral equilibrium model. However, the interpretation of significant results is not always clear. The null hypothesis is a composite hypothesis that includes assumptions regarding the demographics of the populations, such as constant population size and no population structure. There is wide awareness in the field of this fact. For example, when examining the power of the Tajima's D-test, (Simonsen et al. 1995) examined its power against both demographic and selection alternatives. In a simulation study they showed showed that Tajima´s

D reacts to deviations of these assumptions and without additional tests it is not so evident that departure of D=0 is really selection: it could be migration, population growth. The word `neutrality test' has therefore to some degree become synonymous with tests of the equilibrium neutral population model. Significant deviations from the neutral equilibrium model alone do not provide evidence against selective neutrality.

Some insights into the problems associated with these tests have been gained by considering the genealogical structure of the data. For example, a complete selective sweep tends to produce genealogies similar to those generated by a severe bottleneck (Fig. 22). In both cases, the lineages in the genealogy are forced to coalesce at the time of the selective sweep or the bottleneck. The average number of pairwise differences is decreased compared to the number of segregating sites, leading to negative values of Tajima's D. The fundamental problem is that both the demographic process and selection can have very similar effects on the genealogy. It is therefore quite difficult to distinguish these effects when a single locus is considered. For the case of weak selection, it may be even more difficult to use allelic distributions to distinguish selection from demographic processes. (Neuhauser and Krone 1997) have argued that weak selection may at best have only a slight effect on the genealogy. Neutrality tests based on allelic distribution might therefore often have much less power against the common models of selection than against demographic deviations from the neutral equilibrium model.

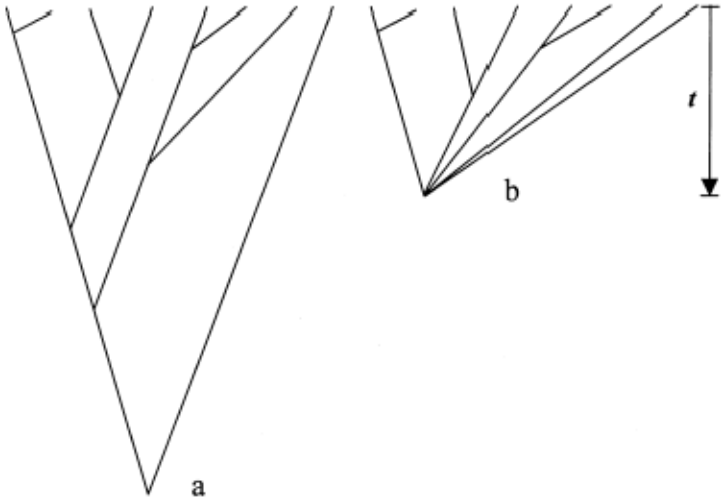


Figure 22. Genealogies simulated under (a) the standard neutral equilibrium model and (b) a model with a severe bottleneck or a complete selective sweep t generations in the past. The effect of a severe bottleneck or a complete selective sweep is to force all lineages in the genealogy to coalesce at the time of the bottleneck/sweep. From (Nielsen 2001).

Another strategy to distinguish selective effects and demography is to compare the result obtained for one gene with a set of 313 genes studied by (Stephens et al. 2001). All the genes have the same demographic history, as it depends on the species, while selection acts differently depending on the gene. Or, in other words, demographic events affect all genes in a genome, whereas natural selection has only local effects. Comparing the value of Tajima´s D of a specific gene with the genes studied by (Stephens et al. 2001) , it is possible to see the position of the gene in relation to the others, and see exactly where in the distribution your gene is located. It is known (Fig. 23) that most of the genes studied give negative values of Tajima´s D, and this is in accordance with the fact that in humans there has been a population expansion or population growth after the *out of Africa* event.

Another study by (Akey et al. 2004) can also be used to compare the values of Tajima´s D in relation with a big sample of genes in order to distinguish demography and selection.

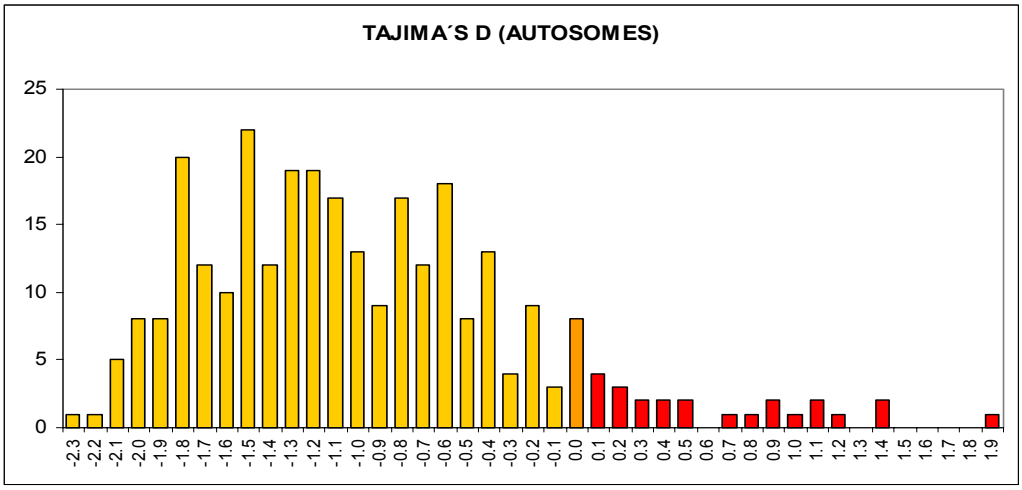


Figure 23. Tajima´s D values from autosomal gene. Different colors distinguish negative values than positive values. Adapted from (Stephens et al. 2001).

In a more recent study, (Stajich and Hahn 2005) concluded that whereas demographic events such as population bottlenecks or expansions will affect all genes in a genome, natural selection is expected to have only locus-specific or region-specific effects on DNA variation. They pointed out that in addition to differences in levels of nucleotide polymorphism, there are also large differences in the average value of Tajima's D statistic (Tajima 1989), a measure of the mutation frequency spectrum, between African-American and European-American populations and from the neutral-equilibrium expectation in both populations. The average D statistic value for the African-American population is -0.49 ($\pm$0.60), and for the European-American population it is -0.26 ($\pm$0.92) (paired t-test, $P < 0.0001$). (Fig. 24).

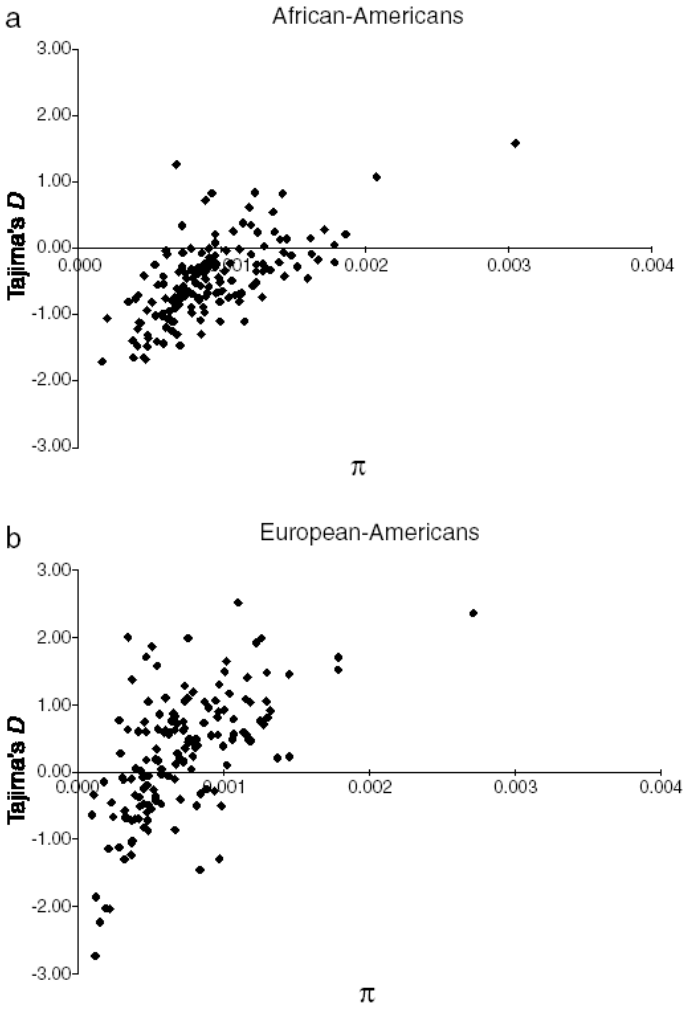


Figure 24. Tajima´s D statistic and pi for all151 loci. (a) African-American population; (b) European-American population.

Their analyses showed that the demographic histories of human populations can largely account for the level and frequency of variation across the genome. However, even working within a nonequilibrium framework, they were able to show deviations from neutral expectations at the ABO and TRPV6 loci and in many regions of low recombination. The results for this data set are consistent with the combined effects of a population bottleneck and repeated selective sweeps in the human migration out of Africa—in agreement with previous

reports (Kayser et al. 2003) (Storz et al. 2004)—and suggest that natural selection affects a relatively large proportion of the genome.

3.2.1.2. Multiple loci

3.2.1.2.1. Lewontin-Krakauer

Several statistical tests have been proposed for employing data from multiple loci. One of the most famous is the Lewontin-Krakauer test (Lewontin and Krakauer 1973). In its original form, this test considers data at diallelic loci from multiple populations. For each locus:

$$F = \sigma_{\mathrm{p}}^2 / [\bar{p}\ (1 - \bar{p})]$$

is calculated, where $\bar{p}$ and σ_p^2 are the mean and variance in allele frequency, respectively, across populations. If the variance in F is too large among loci, the neutral null model can be rejected. The problem with this test is how to determine when the variance in F is too large. In its original form, critical values were calculated assuming independence among populations, a condition that is violated by shared common ancestry or migration between populations (Robertson,1975). The test relies on very strong, and in many cases arguably unrealistic, demographic assumptions.

3.2.1.2.2. HKA

The Hudson-Kreitman-Aguadé (HKA) test (Hudson et al. 1987) compares within-species polymorphism, and between species divergence at two (or more) loci. Under neutrality the level of within-species polymorphism should be correlated with between-species divergence. This degree of correlation should be the same at both loci if they are evolving in a neutral fashion. Thus the null hypothesis of neutrality can be tested at a chosen locus by comparing it with sequence polymorphism and divergence at a neutral control locus in the same two species.

Under neutrality, three parameters are expected to relate observed diversity within and between the two species: θ; the time since divergence; and the ratio of effective population sizes. In essence, these are estimated from the observed diversity at all loci, and are then used to generate expected polymorphism and divergence for each locus. A statistical method known as *goodness-of-fit test* is then used to compare these locus-specific measures of observed and expected diversity. If neutral evolution is operating at both loci, the expected and observed measures of diversity agree reasonably well. Selection is inferred when the variance among loci of the ratio of divergence to polymorphism is too high.

While originally formulated for comparing two unlinked loci, this test has been further developed to compare linked regions within the same gene. A more recent adaptation slides a window along a sequence, to explore selective differences within a single gene. This type of analysis can reveal situations where one protein domain, but not others, has been under selection.

3.2.2. Comparing variability in different classes of mutations

3.2.2.1. McDonald and Kreitman test

Tests based on allelic distribution or variability alone are quite sensitive to the underlying demographic assumptions, mostly because the structure of the gene genealogy is a product of the demographic processes in the populations. However, it is possible to establish tests of neutrality based on statistics with distributions that are independent of the genealogy or only depend on the genealogy through a *nuisance parameter* that can be eliminated. A famous example is the McDonald-Kreitman test (McDonald and Kreitman 1991). In this test, the ratio of nonsynonymous to synonymous polymorphisms within species is compared to the ratio of the number of nonsynonymous and synonymous fixed differences between species in a 2 x 2 contingency table. The justification of this test is very similar to the HKA test. The test is based on the following assumptions:

1- Only nonsynonymous mutations may be adaptive
2- Synonymous mutations are always neutral, and
3- Selectively advantageous mutations will be fixed in the population much more rapidly than a neutral mutation, and hence are less likely to be found in a polymorphic state.

If polymorphism and divergence are driven only by mutation and genetic drift, the ratio of the number of fixations to polymorphisms should be the same for both nonsynonymous and synonymous mutations. A significant difference

between the two ratios can therefore be used to reject the neutral mutation hypothesis.

In statistics, parameters that are of no interest to the researcher but cannot be ignored are labelled *nuisance parameters*. A common approach is to eliminate such parameters by conditioning on a sufficient statistic, i.e. a statistic that contains all the relevant information to the data regarding the parameter. In the case of the McDonald-Kreitman test, the total tree length is the *nuisance parameter* and the total number of substitutions is a sufficient statistic for this parameter. By conditioning on the total number of substitutions in the 2 x 2 table, the total tree length parameter is eliminated. In this manner a test of neutrality is established that is valid for any possible demographic model. The McDonald-Kreitman test has been very useful for detecting selection. For example, (Eanes et al. 1993) found very strong evidence for selection in the G6pd gene in *Drosophila melanogaster* and *D. simulans*. Although the McDonald-Kreitman test does provide unambiguous evidence for selection, it is not always clear which type of selection is acting on the gene. For example, changes in the population size combined with weak selection against slightly deleterious mutations may either increase or decrease the number of nonsynonymous polymorphisms. An increase in the population size will lead to a deficiency of nonsynonymous polymorphisms and a decrease in population size will lead to an excess of nonsynonymous polymorphisms. Significant results from the McDonald-Kreitman cannot be interpreted directly as evidence for positive selection.

This test does not rely on assumptions regarding the demographics of the populations because it is constructed by comparing different types of variability within the same locus, or genomic region. Since nonsynonymous and synonymous sites, for example, are interspersed among each other in a coding region, the effect of the demographic model is the same for both types of site.

3.2.2.2. Tests based on the dN/dS ratio

In the coding region we can observe two types of nucleotides, the ones that when mutated result in a different aminoacid (or non-synonymous sites) or when mutated result in exactly the same aminoacid as before (synonymous sites). These last type of change are considered selectively neutral, as they should not be affected by natural selection and so the proportion of such sites (dS or Ks) may be independent of this force. The proportion of nonsynonymous sites that are variable within the same set of sequences (dN or Ka) will be greater than dS under diversifying selection, and less than dS under purifying selection. This ratio (Ka/Ks) is often known as ω. By testing if ω is significantly different from 1, we are in effect testing the difference between dN and dS. If ω is significantly greater than 1, diversifying selection would appear to be acting. If ω is significantly less than 1 (in most of the genes), then purifying selection predominates.

This approach presents lack of statistical power when this is applied to very close species because of the little information in the data and also an underestimation of reversions between sequences are observed when very

distant species are considered. This underestimation of distances can be corrected by using a model of sequence evolution (Yang and Bielawski 2000).

A statistical framework for making inferences regarding dN and dS was developed by (Goldman and Yang 1994) and (Muse and Gaut 1994). In this framework the evolution of a nucleotide sequence is modelled as a continuos-time Markov chain with state space on the 61 possible codons in the universal genetic code. In one parameterization, the instantaneous rate matrix of the process Q , is given by:

$$q_{ij} = \begin{cases} 0, \text{ if the two codons differ at } > 1 \text{ position}, \\ \pi_j, \text{ for synonymous transversion}, \\ \kappa\pi_j, \text{ for synonymous transition}, \\ \omega\pi_j, \text{ for nonsynonymous transversion}, \\ \omega\kappa\pi_j, \text{ for nonsynonymous transition} \end{cases}$$

where π_j is the stationary frequency of codon j, κ is the transition/transversion rate ratio and ω (=dN/dS) is the nonsynonymous/synonymous rate ratio. Using this model, it is possible to calculate the likelihood function for ω and for other parameters using the general algorithm of (Felsenstein 1981). It is thereby possible to obtain maximum likelihood estimates of these parameters, and hypotheses such as H_0: $\omega \leq 1$ can be tested using likelihood ratio tests. This maximum likelihood method has several advantages over previous methods in that it correctly accounts for the structure of the genetic code, it can incorporate complex mutational models and it is applicable directly to multiple

sequences, taking the structure of the underlying genealogical tree into account.

Using the synonymous substitution rate as a proxy for neutral evolution avoids possible problems associated with differential mutation rates at different loci (caused by differences in sequence composition or genealogical history), but further complication arises from the fact that many organisms show biases in their codon usage resulting from both mutational bias and selection for the efficient translation of proteins. As a consequence, evolution at synonymous sites may not be truly neutral.

As described before, ω is the average over many nucleotides. However, selective pressures are likely to be different among nucleotides within the same gene (some sites might be under negative selection while mutations at others may be neutral or even undergo positive selection), and thus meaningless. However, when many sequences can be compared, it does become possible to detect selective pressures at individual sites using codon-based tests, and ω can also be used to compare different portions of the same gene to detect whether certain functional modules have been under different selective regimes.

The comparative analysis of whole genome sequences provides an opportunity to use ω to detect selection patterns over a huge number of genes. An analysis of 12.845 genes orthologous between mouse and human (Mouse Genome

Sequencing Consortium 2002) showed that the median value of ω among these genes was 0.115, indicating that most are under purifying selection.

As this purifying selection is expected as genes need a high degree of conservation to maintain their function, normally dN/dS ratio is calculated to see if there is positive selection. Only very low levels of ω are take into account to observe specific and strong purifying selection in a gene.

A list of genes that have been shown to be under positive selection in primates from a consideration of ω are shown (Fig. 25) (Vallender and Lahn 2004).

Category	Gene	Lineage
Host–pathogen interations	ABO blood group	Human/Primates
	APOBEC3G	Primates
	Beta globin (*HBB*)	Recent human
	Beta-defensin	Primates
	CD45	Catarrhines
	CD59	Humans
	Chemokine receptor 5 (*CCR5*)	Human
	Cytidine monophospho-*N*-acetylneuraminic acid hydroxylase (*CMAH*)	Human
	Duffy blood group	Recent human
	Eosinophil cationic protein (*ECP*)	Catarrhines
	Glucose-6-phosphate dehydrogenase (*G6PD*)	Recent human
	Glycophorin A	Recent human/Primates
	Immunoglobulin A	Primates
	Immunoglobulin heavy chain	Mammals
	Interleukin 2 (*IL2*)	Mammals
	Interleukin 4 (*IL4*)	Human
	Killer cell inhibitory receptors (*KIR*)	Human
	Major histocompatibility complex (*MHC*)	Mammals
	Pyrin	Recent human
	Rh blood group	Hominids
	Sialic acid-binding immunoglobulin-like lectin 9 (*SIGLEC9*)	Human
	Sialic acid-binding immunoglobulin-like lectin-like protein 1 (*SIGLECL1*)	Human
	Tumor necrosis factor ligand superfamily, member 5 (*TNFSF5*)	Recent human
Reproduction	Chorionic gonadotropin	Primates
	Chromodomain protein Y (*CDY*)	Primates
	Deleted in azoospermia (*DAZ*)	Primates
	Fertilin	Primates
	Oviductal glycoprotein 1 (*OVGP1*)	Mammals
	Protamine 1 (*PRM1*)	Primates/Mammals
	Protamine 2 (*PRM2*)	Primates/Mammals
	Semenogelin 1 and 2 (*SEMG1*, *SEMG2*)	Primates
	Sex-determining region Y (*SRY*)	Primates/Mammals
	Sperm associated cation channel 1 (*CATSPER1*)	Primates
	Sperm protein associated with the nucleus, X-chromosome (*SPANX*)	Hominids
	Transition protein 2 (*TNP2*)	Primates/Mammals
	Zona pellucida glycoprotein 2 (*ZP2*)	Mammals
	Zona pellucida glycoprotein 3 (*ZP3*)	Mammals
Dietary adaptation	Alanine:glyoxylate aminotransferase	Anthropoids
	Aldehyde dehydrogenase 2 (*ALDH2*)	Recent human
	Lactase (*LCT*)	Recent human
	Lysozyme (*LYZ*)	Human/Primates
Appearance	Melanocortin 1 receptor (*MC1R*)	Recent human/Mammals
Sensory systems	MAS-related genes (*MRG*)	Primates
	Olfactory receptors (*OR*)	Human/Primates
	Red and green opsins	Human/Primates
	Taste receptor type 1 member 3 (*TAS1R3*)	Primates
	Taste receptor type 2 (*TAS2R*)	Primates
	Taste receptor type 2 member 38 (*TAS2R38*)	Recent human
	Type 1 vomeronasal receptor-like (*V1RL*)	Primates
Behavior	Dopamine receptor D4 (*DRD4*)	Primates
	Forkhead box P2 (*FOXP2*)	Human
	Monoamine oxidase A (*MAOA*)	Human
Brain anatomy	Abnormal spindle-like microcephaly associated (*ASPM*)	Human
	Microcephalin (*MCPH1*)	Human
	Myosin heavy chain 16 (*MYH16*)	Hominine
Miscellaneous	Angiogenin (*ANG*)	Primates
	Breast cancer 1 (*BRCA1*)	Human
	Cytochrome *c* oxidase subunit 4 (*COX4*)	Human/Primates
	Cytochrome *c* oxidase subunit 7a (*COX7A*)	Primates
	Cytochrome *c* oxidase subunit 7c (*COX7C*)	Primates
	Cytochrome *c* oxidase subunit 8 (*COX8*)	Human/Primates
	Forkhead box D4 (*FOXD4*)	Hominids
	Morpheus	Hominids

Table 5. Genes that have been shown to be under positive selection in primates. From (Vallender and Lahn 2004)

3.2.3. Tests of neutrality based on SNP data

Other strategies based on SNP data to detect selection have been developed recently. It is important to point out that all of the tests described before had been based on data obtained by sequencing, and this meant that all the polymorphisms, from high frequency SNPs to singletons, were represented.

Linkage disequilibrium (LD) across the human genome offers a new way to detect a signature of recent positive selection (Sabeti et al. 2002) (Nordborg and Tavare 2002). The logic underlying this strategy is straightforward. When a mutation arises, it does so on an existing background haplotype characterized by complete LD between the new mutation and the linked polymorphisms (Fig. 25).

Over time, new mutations and recombination reduce the size of this haplotype block such that, on average, older and relatively common mutations will be found on smaller haplotype blocks (that is, there is only short range LD between the mutation and linked polymorphisms).

Younger, low-frequency mutations might be associated with either small or large haplotype blocks. A signature of positive selection is indicated by an allele with unusually long-range LD and high population frequency. The formal implementation of this strategy has recently been introduced as the long range haplotype (LRH) test (Sabeti et al. 2002).

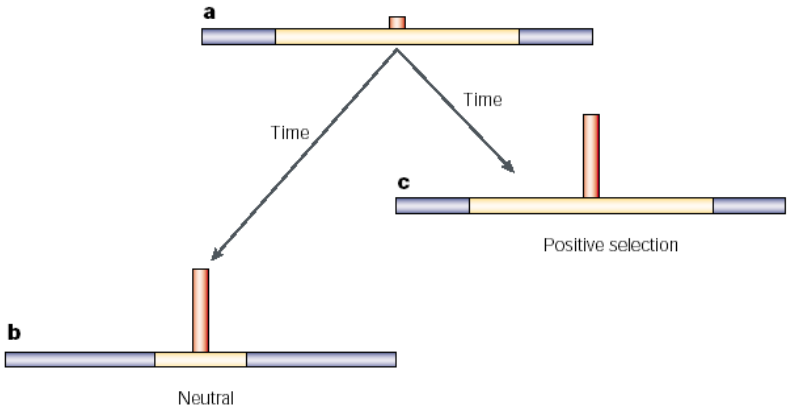


Figure 25. Detecting recent positive selection using linkage disequilibrium analysis. a) A new allele (red) exists at a relatively low frequency (indicated by the height of the red bar) on a background haplotype (blue) that is characterized by long-rage linkage disequilibrium (LD) (yellow) between the allele and the linked markers. b) Over time, the frequency of the allele increases as a result of genetic drift, and local recombination reduces the range of the LD between the allele and the linked markers (that is, it creates short-range LD). c) An allele influenced by recent positive selection might increase in frequency faster than local recombination can reduce the range of LD between the allele and the linked markers. From (Bamshad and Wooding 2003).

4. OBJECTIVES

The general objective of this thesis is the comprehension of the diversity and evolution in a particular gene, PRNP, involved in prion diseases, through a comparative study of different human populations and in the closest species, chimpanzees.

The specific objectives in each section that constitute the global project are the following:

PRION PROTEIN GENE IN HUMANS

The general objective of this section is the understanding of the evolution of this gene, PRNP, in the human species. Analysing the prion gene in different populations representing the genetic variation in humans is necessary to detect, if it is the case, the differences between human groups.

In the first place, the aim is to analyse specific positions in the gene that had been previously described as being related with disease (being causative or increasing the susceptibility or protecting from prion diseases). It is of interest to see if the distribution of these polymorphisms is significantly different among human populations, a fact that can be related to different susceptibility.

Another objective was based on designing a new methodology where several of the important SNPs could be studied at the same time.

For these reasons we decided to genotype 5 positions in the gene that were relevant in human prion diseases (102, 129, 178, 200 and 219) in a set of human samples representing the genetic variation of our species (CEPH-HGDP

diversity panel). We include all the positions in a methodology that with a single procedure, all the SNPs were genotyped.

The main conclusions of this study can be found in **Chapter II** of the **Results section**.

A second objective was to analyse the sequence diversity and infer the evolution of the gene PRNP, and detect specifically if this locus has been under selective forces or under neutrality. If selective forces were acting in this gene, determine what type of selection could be acting.

For obtaining this information we decided to sequence the complete exon 2 region, where the coding part is located, in 348 human chromosomes from populations representing the genetic variation of our species.

Conclusions of this study can be found in **Chapter IV** of the **Results** section.

As a similar study was concurrently being carried out by a competing research group but based on another approach, comparison of both studies and conclusions can be found in **Chapter III** of the **Results** section, where the main results of their study are challenged.

In the third place, our interest was also to focus on the analysis of prion gene variation in patients with prion diseases. Little is known about the variation outside the coding region of the gene in patients suffering from prion diseases. Moreover, the high ratio of patients with sporadic CJD found are still not understood, therefore perhaps more detailed analysis of the variation of other

parts of the gene may generate new data leading to greater understanding of prion diseases.

For these reasons we sequenced the promoter region and the complete exon 2 of the PRNP in a total of 119 patients (familial and sporadic). Conclusions of this study can be found in **Chapter V** of the **Results** section.

PRION PROTEIN GENE IN CHIMPANZEES

The general objective of this section is the understanding of the differences between humans and chimpanzees regarding the PRNP gene. We were particularly interested in the well-known polymorphism in codon 129 of this prion gene that has been found to be important in prion diseases susceptibility. At the same time the analysis of this gene in these two species can provide knowledge about the differences between these two species. There is an increasing interest in studies of comparative genetics between humans and chimpanzees, and this can help to look for species-specific genes. Moreover, it can help to explain why one species is more susceptible to prion diseases than another.

These types of studies, even if they provide much information, are limited in number because of the increasing difficulties to obtain a good and large sample of non-human primates.

Thanks to other groups and the collaboration of zoological parks it was possible to work with 65 samples of *Pan troglodytes* that allowed us to perform an

intraespecific study of this species in the PRNP gene. Conclusions can be found in **Chapter I** of **Results** section.

5. RESULTS

5.1. CHAPTER I

Soldevila M, Andrés AM, Blancher A, Calafell F, Ordoñez M, Pumarola M, Oliva B, Aramburu J, Bertranpetit J.
Variation of the prion gene in chimpanzees and its implication for prion diseases.
Neurosci Lett. 2004 Jan 30;355(3):157-60.

5.2 CHAPTER II

Soldevila M, Calafell F, Andrés AM, Yagüe J, Helgason A, Stefánsson K,Bertranpetit J.
Prion susceptibility and protective alleles exhibit marked geographic differences.
Hum Mutat. 2003 Jul;22(1):104-5.

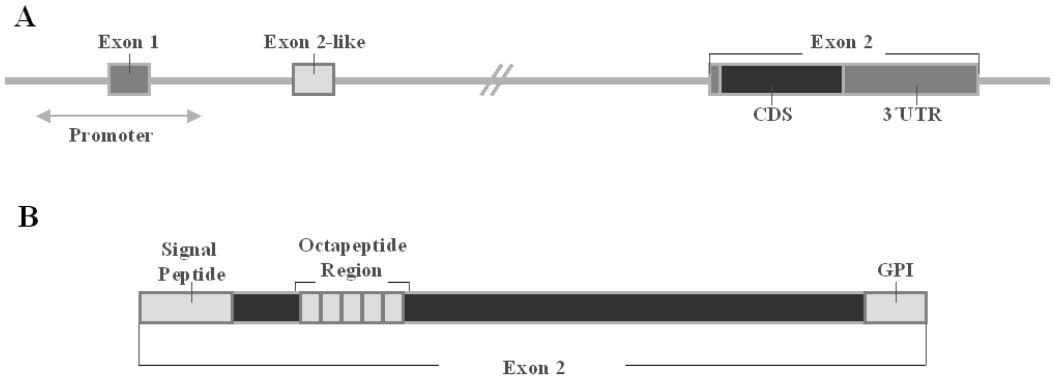
5.3

General overview of the samples used and regions analysed in some of the following chapters.

Prion protein (*PRNP*) gene

The *PRNP* gene coding for the prion protein has been mapped to chromosome 20p12-pter.

The general structure of the *PRNP* gene is the following:



In humans this gene is 15.185 bp long, and it contains 2 known PrP exons (134 and 2.355 nucleotides in length) separated by an intron of 12.696 bases. Genomic sequencing allowed the identification of an additional exon-like sequence analogous to exon 2 of sheep and mouse, but any *PRNP* transcript does not include the untranslated exon 2 found in these organisms.

In humans the *PRNP* gene encodes the 253-amino-acid PrP (CDS is in the exon 2).

Moreover, an increasing number of point and insert mutations in *PRNP* have been associated to familial forms of spongiform encephalopathy: familial CJD, GSS and FFI.

Specific proposal

To determine the sequence of the following regions:

Region 1 and region 2 will be about 3 kb in length each. We will include all three exons (one exon-like) and the CDS (739 bp) that is part of exon 2.

Primers used for each region:

REGION 1	5´->3´
DP1F	CAAGAACACACCACCCTGT
DP1R	CCGGCCGAGGTTTAAGTT
UP3F	TGCTTCCTCATTCCTGAGCC
UP3R	TTTACCTGCCTCGGTCGTG
UPSEQF	CATTCCTGAGCCTTTCATTTTC
INT1F	TTCTCCTCTCCTCACGACCG
INT1R	TCTCCATCCCCCCAAGC
INTSEQF	TCACGACCGAGGCAGGTAAAC
int2F	GATGCCCAGAGGGTGCTT
int2R	GTGACCCAAGGGTCGTCAT
DMAB-5F	AGGTGCGATGGCGGACCCCAGC
DMAB-8R	ACTGTAAGACCTTCTCTGTG
DMAB-7F	TTGTTCTGAGAGCATCACGG
DMAB-10R	TTCCAGCTTCTTAATGCATCAC
DMAB-9F	TCACAAAGATGGTTCTGCAGTC
DMAB-11F	ATGATTTTGGAGCCTGTGATATG
DMLBbb	TGGCCTCTGGATCAAGAGCTTGTG
DMAB-12R	AACTGTGGGTCCATTTCATC

REGION 2	5´->3´
-21	GTTCACCCTTTTCTTACTTTTG
-21	TGATGGGCCTGCTCATGGCAC
PRION HX-A	TGTAAAACGACGGCCAGTGCAGTCATTATGGCGAACC
PRION HX-B	CAGGAAACAGCTATGACCGGCA CTTCCCAGCATGTAGCCGC
PRION HX-C	TGTAAAACGACGGCCAGT TGGCACCCACAGTCAGTGGAA
PRION HX-D	CAGGAAACAGCTATGACC CCTTCCTCATCCCACTATC
DMCD-3F	ATCAAGCAGCACACGGTCACC
DMCD-5F	AGTCTGAAATACCTTTGCCTGG
DMCD-6R	AGCCAGAGGTTCAGTGTTGTGAC
DMCD-7F	TCTCCTTTGTCCATTTACCTGG
DMCD-8R	ACAATTCAGGGAATAATTTTAC
DMCD-10R	CGGGAGGTGGTGCTCATCTTC

Samples to be sequenced:

1. 348 human chromosomes from a worldwide distribution (HGDP-CEPH panel):
    - AFRICANS- Mbuti Pygmy, Biaka Pygmy, San, Mandenka and Yoruba.
    - E-ASIA: Japan, Han, Yakut and Cambodia
    - W-ASIA: Druze, Bedouin, Palestinian and Sindhi.
    - EUROPEANS: Orcadian, Sards, Russian, Basques and French.
    - AMERICANS: Maya
    - PACIFIC: Nan melanesian and Pacific.
2. Icelandic families (96 individuals).
3. primates:
    1. 30 chimpanzees
    2. And a few gorilla samples in order to generate external references

4. 119 Creutzfeldt-Jakob patients (sporadic and familial)

1. Screening of SNPs of interest in a much larger human sample or in order to be used in case-control studies.

   -For polymorphisms in codon 129 and 219: The complete HGDP-CEPH panel (1064 samples).

   -For polymorphism in codon 129: 96 extra samples of Icelandic population.

   -About 2 x 96 control samples (including Catalan and Basque populations) in order to be used as a control for CJD patients. (SNPs -563, -101, +310 in the promoter region).

Initial specific goals

- Generate a detailed analysis of their variation of the PrP region in humans, unravelling new variants and geographic distribution.
- Recognize the effect of selection along the analysed genome region.
- Understand the time and space framework of the known variants of resistance to prionic disease.
- Comparative analysis of humans and chimpanzees. Differences will be possible to be attributed to selective forces or to the species dynamics (neutral variation).
- Susceptibility alleles (such as Methionine in codon 129): it will be possible to recognize when they appeared, whether they have a unique origin and whether natural selection has acted to increase or reduce its frequency.

5.3.1 CHAPTER III

Soldevila M, Calafell F, Helgason A, Stefánsson K, Bertranpetit J. *Assessing the signatures of selection in PRNP from polymorphism data: results support Kreitman and Di Rienzo's opinion.* Trends Genet. 2005 Jul;21(7):389-91.

5.3.2 CHAPTER IV

M. Soldevila, A.M. Andrés, A. Helgason, A. Ramírez-Soriano, T. Marquès-Bonet, S. Sigurdadóttir, F. Calafell, A. Navarro, K. Stefansson and J Bertranpetit.

The prion protein gene in humans revisited: lessons from a worldwide resequencing study

***In preparation*.**

THE PRION PROTEIN GENE IN HUMANS REVISITED: LESSONS FROM A WORLDWIDE RESEQUENCING STUDY

M. Soldevila[1], A.M. Andrés[1*], A. Helgason[2], A. Ramírez-Soriano[1], T. Marquès-Bonet[1], S. Sigurdadóttir[2], F. Calafell[1], A. Navarro[1], K. Stefánsson[2] and J Bertranpetit[1].

1. Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra

   Address: Dr. Aiguader 80, 08003 Barcelona, Spain

   Telephone number: (34) 93 542 28 39 *Fax number*: (34) 93 542 28 02

2. deCode genetics

   Address: Sturlugata 8, IS-101 Reykjavik, Iceland .

   Telephone number: (354) 570 19 00 *Fax number*: (354) 570 19 03

   (*) current address: Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA

Corresponding author: Jaume Bertranpetit

e-mail: jaume.bertranpetit@upf.edu

Running title: Haplotype variation in exon 2 of prion protein gene

RESULTS

ABSTRACT

There is strong evidence to suggest that the PRNP gene is involved in prion diseases. In particular, variation in codon 129 appears to be a strong modifier of risk in human prion diseases such as Creutzfeld-Jakob disease (CJD). Here we evaluate the patterns of nucleotide variation in PRNP exon 2, which includes all the coding sequence, by sequencing a worldwide sample of 174 humans. In line with previous studies, we found two main haplotypes, differentiated by the non-synonymous substitution in codon 129, that accounted for most of the overall diversity found in this exon. Our analyses reveal the worldwide pattern of variation at the PRNP gene to be inconsistent with neutral expectations, indicating instead an excess of low-frequency variants that is more consistent with the action of either positive or purifying selection. A comparison of selection test statistics for PRNP with those obtained from sequence data for over 292 human genes indicates that the signal of positive selection on PRNP is stronger than expected from a possible confounding genomewide background signal of population expansion. Two main conclusions arise from our analysis. First, the existence of an ancient, estable, balanced polymorphism that has been claimed in a previous study (Mead et al. 2003) can be discarded. Second, our results are consistent with a complex history of episodic or fluctuating selection; most likely short periods of positive selection with the associated selective sweeps, followed by drift and/or purifying selection. It is alsopossible that very short periods of balancing selection produce molecular signatures similar to a partial selective sweep, so episodes such as Kuru are also consistent with our data.

INTRODUCTION

Transmissible spongiform encephalopathies (TSEs), or prion diseases, are a group of rare, sub-acute and fatal neuro-degenerative disorders characterized by accumulation of the abnormal isoform of a host-encoded membrane protein. Human TSEs, can be sporadic, acquired, or inherited, and include Creutzfeldt–Jakob disease (CJD), Gerstmann-Straüssler-Scheinker (GSS) disease, fatal familial insomnia (FFI), kuru (a disease confined to a population in Papua-New Guinea) and variant CJD (vCJD), a concept coined to designate cases potentially caused by the human consumption of cattle suffering from bovine spongiform encephalopathy (BSE).

The human prion protein (PrP) is a product of a single gene located on the short arm of chromosome 20 (Prusiner 1991a).

The prion protein is encoded by a single exon of the *PRNP,* exon 2 (referred to as exon 3 in some publications). Only part of this exon is translated into the protein, the rest of exon 2 and the whole of exon 1 remain only transcribed. Variation in the gene sequence produce protein variants that are causative of inherited TSE diseases (the most common are at codons 200, 178 and 102)Other polymorphisms have been linked to differential susceptibility to the acquired TSE diseases, with those in codons 129 and 219 being the most important (refs?). In the case of codon 219, it has been suggested that the Lys allele acts as a protective factor against sporadic CJD. Thus, all Japanese individuals with sporadic CJD were 219Glu homozygotes, while this polymorphism was present at 6% frequency in Japanese controls (Shibuya et al. 1998). It was later shown that the protective 219L allele is restricted to Asian and Pacific populations (Soldevila et al. 2003).

RESULTS

The common methionine/valine (M/V) polymorphism at codon 129 is generally considered to be the most important in the genetics of prion diseases. Previous studies have indicated that this polymorphism modifies susceptibility to CJD. Up to 90% of sporadic CJD (sCJD) cases have occurred in individuals that are homozygous for either versions of codon 129, MM or VV (Palmer et al. 1991). So far all the vCJD cases reported have been MM homozygotes (Valleron et al. 2001); (Andrews et al. 2003); (Collinge et al. 1996). Furthermore, this polymorphism of the *PRNP* gene has an influence both on the disease phenotypes of sCJD (Parchi et al. 1999) and on the disease resulting from a mutation at codon 178 (Goldfarb et al. 1992).

Codon 129 is also a key factor in the resistance or susceptibility to the kuru prion disease, which was thought to have been transmitted during endocannibalistic feasts among the Fore linguistic group in New Guinea. (Cervenakova et al. 1998) analysed a group of Fore individuals affected with kuru and observed that homozygotes had an earlier age at onset (and thus probably a shorter incubation period) than heterozygotes.

A similar approach was taken by (Mead et al. 2003), who screened Fore women over fifty years of age, many of whose cohort would have died of kuru. Among these women, there was a statistically significant excess of heterozygotes for the codon 129 polymorphism, implying a heterozygote resistance to the disease. Based on an analysis of *PRNP* haplotype diversity in a worldwide sample, Mead et al. (2003) postulated that this locus had been affected by strong balancing selection during the evolution of modern humans. This conclusion was mainly founded on the apparent predominance of intermediate polymorphisms in their data set, a pattern of variation that is consistent with the impact of balancing selection. However, as pointed out by (Kreitman and Di Rienzo 2004) the study of Mead et al. (2003) was affected by a bias in the ascertainment of genetic variants. Thus, Mead et al. (2003) first screened *PRNP* polymorphisms by

sequencing a limited number of individuals, and then typed the polymorphisms thus identified in a larger worldwide sample of individuals. This led to the exclusion of all low frequency variants with a 5% frequency cut-off and thereby artificially tilted the resulting allele-frequency spectrum towards a pattern consistent with balancing selection, summarized as strongly positive values by statistics such as Tajima's D.

At the same time, in another crucial species, *Bos taurus*, purifying selection was found to be the main force driving the evolution of *PRNP* gene (Seabury et al. 2004). In this study exon 3 (equivalent of exon 2 in humans) was resequenced in 36 breeds of domestic cattle obtaining large negative values of Tajima´s D, and when comparing the ratio of non-synonymous and synonymous variants (dn/ds) with other species they observed values much lower than 1, indicating that purifying selection was acting in this gene both at an intraspecific level and at the interespecific level.

A worldwide survey of codon 129 revealed marked geographic differences in frequencies of the two variants among continents (Soldevila et al. 2003). The 129M susceptibility allele was found at high frequencies in Old World populations, very high in the Pacific (~81%) and up to 93% in Central and East Asia, but at a low frequency (~30%) in Native Americans.

Thus, different types of selection have been postulated in this gene in different species and doubts exist on the adequacy of data for humans. In order to clarify the selective pressures that have acted on the *PRNP* gene in humans, we decided to resequence the entire exon 2 of *PRNP* in a sample of 348 chromosomes of humans from worldwide populations.

RESULTS

MATERIALS AND METHODS

Samples

The HGDP-CEPH Human Genome Diversity Cell Line panel contains a total of 1064 samples from a broad range of different world populations (Cann et al. 2002). From this panel, we selected a total of 174 individuals (348 chromosomes) for sequencing PRNP exon 2, while the entire panel was used for genotyping the SNP polymorphisms in codons 129 and 219. For sequencing, samples were chosen to represent the following large areas and populations: Africa A (2N=28 from Mbuti and Biaka Pygmy groups), Africa B (2N=40 from San, Mandenka, and Yoruba), Europe (2N=104 from Basques, Orcadian, Sardinian, Russian, and French), Pacific (2N=32 Noan-Austronesian speakers in Papua New Guinea), West Asia (2N=64 from Druze, Bedouin, Palestinian, and Sindhi), East Asia (2N=64 from Japan, Han, Yakut, and Cambodia) and Native Americans (2N=16 from Maya). One chimpanzee (from the Barcelona Zoo) was sequenced in order to assess the ancestral state of alleles.

SNP genotyping

Codon 129 and 219 were genotyped using *TaqMan* technology from Applied Biosystems (AB) for the entire HGDP-CEPH diversity panel. The Assays-on-Demand™ service was used to design probes and primers for codon 129 and the Assays-by-Design℠ service was used to design probes and primers for codon 219. Assays-on-Demand™ SNP Genotyping products consist of a 20X mix of unlabeled PCR primers and TaqMan® MGB (minor groove binder) probe (FAM™ and VIC® dye-labeled). Assays-by-Design℠ service consists on a 40X mix of unlabeled PCR primers and TaqMan® MGB probe (FAM™ and VIC® dye-labeled). All assays were

optimized to work with TaqMan Universal PCR Master Mix and with genomic DNA (10 ng). The total volume used was 5 μl/well (384 well-plate). Amplification conditions were as follows: 50ºC, 2 min; 95ªC, 10 min; followed by 40 cycles of 94ªC, 15 sec and 60ªC, 1 min, in ABI Prism 7900 HT (AB). . Fluorescence in each well was measured after PCR and the results were analysed with SDS software package version 2.1 (AB). Success rate was 99.25% for codon 129 and 98.5% for codon 219 (Table 1).

Sequencing

All site positions referred to in this study are in accordance with a reference sequence from (Lee et al. 1998), Genbank accession number U29185. Note that there is a deletion in the octapeptide repeat region (R3-R4, deletion C) in the reference sequence, so that the region amplified is usually 24 bases longer , for a total of 2,378 bp. Amplification reactions of exon 2 of PRNP gene were performed on 10 ng of template DNA in a 20 μl volume by use of Ampli*Taq* Gold polymerase (PE Biosystems). PCR primers and conditions are available on request. Segments were sequenced by use of the BigDye Terminator Cycle Sequencing kit from PE Biosystems on an ABI 3700 (AB) DNA sequencer. The amplification primers were used for cycle-sequencing reactions. All samples were sequenced for both strands, providing an overlap between the PCR segments. The sequences were aligned with SEQMAN II 4.03 (DNASTAR) and manually checked. The latter step turned out to be very important as X (6%) of the Y polymorphic sites reported in this study would not have been detected based only on automatic allele calling software.

Cloning

RESULTS

PCR products with sequencing problems because of insertions or deletions (indels) were cloned with pMOSBlue blunt ended cloning kit (Amersham Biosciences) following the manufacturer's instructions. A total of 7 μl of PCR product were treated with pK enzyme mix, incubated at 22ºC for 40 minutes and ligated into pMOSBlue vector overnight. 1.5 ml of the ligation product were transformed into 20 ml of competent cells, grown in 80 ml of SOC medium at 37ºC for one hour and plated on IPTG/X-gal agar plates. After 16 hours, white colonies were subjected to direct PCR screening using the same primers used in the PCR amplification. Length polymorphisms were identified by agarose gel electrophoresis, purified and sequenced. This procedure allowed the ascertainment of the length and composition of each haploid sequence genotype.

Data analysis

Haplotype inferences and median-joining networks

Haplotype frequencies were estimated using the Bayesian approach implemented in the Phase 2.0 software The Network 4.1.0.0. software was used to generate median-joining networks describing possible genealogical relationships among haplotypes in terms of mutational differences (Bandelt et al. 1995). The low variation detected in the PRNP sequences enabled made the determination of phase relatively straightforward and simplified the resolution of reticulations in the network.

Neutrality tests and diversity statistics

The following test statistics to detect signals of natural selection, Tajima´s D, Fu and Li´s F*, Fu and Li´s D*, Fay and Wu´s H tests, MK test, and various diversity statistics were calculated using DnaSP 4.00

The recombination rate was estimated from (Kong et al. 2002), where the closest STR was D20S97, with an estimated sex-averaged recombination rate of 2.94 centiMorgans per megabase (cM/Mb). For the studied length of 2,378 bp, the recombination rate per generation is $7.0x10^{-5}$. From this, the obtained R value was 2.79 (R= 4Ner), where r is the recombination rate per generation and Ne the effective population size (10,000). The distribution of the observed pairwise nucleotide site differences and the expected values (for no recombination) in growing and declining populations were obtained using the programs Arlequin (Schneider et al 2000) and Network 4.0 (Bandelt et al. 1995).

TRMCA and age of mutations

The Genetree software was used to estimate the coalescence times and the ages of mutations (Tavare et al. 1997), which provides an efficient method for generating likelihood surfaces for $\theta = 4N_e \mu$ given a tree and the frequencies of the variants. These estimates were computed both assuming a constant population and growing population sizes

Substitution Rate

We have estimated the substitution rate in exon 2 as the number of differences over 2tL, L being the length of the segment compared and t the divergence time between species. With 26 differences with chimpanzee, divergence is 1.09%, a similar value to many others published in recent literature. If a divergence time of 6 million years is assumed between humans and chimpanzees, the average substitution rate is 0.91×10^{-9} per nucleotide and year for the studied region.

Genetic structure statistics

RESULTS

F_{ST} values and AMOVA were calculated using the Arlequin software (Schneider et al. 2000). These analyses were performed for the sequence data (2N=348) and also the data for codons 129 and 219 (2N = 2128) obtained using the TaqMan assays.

RESULTS

Codons 129 and 219

Typing the worldwide CEPH-HGDP panel for codons 129 and 219 confirmed and expanded the results obtained in a smaller previous study (Soldevila et al. 2003). For codon 129 there is considerable geographic heterogeneity (Table 1), with frequencies of the valine allele ranging, at the continent level, from 5% in East Asia to 65% in the Americas. We note that the high frequency of 129Val in the Americas is consistent among the five subpopulations examined in this continent, albeit with a slightly higher frequency in South America.

Codon 219 was found to be monomorphic in Europe, Africa and Americas, with the 219Lys allele observed only in Asian and Pacific populations, and very rare in the Middle-East/North African population (Table 1). Two homozygotes for 219Lys were detected, both from Central-South Asian populations (Pathan and Sindhi, both in Pakistan) and were confirmed by sequencing. Hardy-Weinberg equilibrium holds for all populations (a single case had a p=0.019, non-significant after Bonferroni correction for multiple testing).

PRNP Nucleotide Sequence Variation

For the 2,378 bp sequenced in exon 2 of the *PRNP* gene (Fig. 1) in 174 humans, 22 variants were observed, thereof 18 SNPs and 5 length polymorphisms. Of the latter, three were identified in the octapeptide repeat region. This region consists of a series of repeats, usually five, named R1 to R4 depending on their sequence, where each codes for eight amino-acids, with the exception of R1, which codes for nine. The composition of this region is generally known as R1-R2-R2-R3-R4, which reflects the organization

of the reference sequence. We observed one chromosome with a deletion of an R2 repeat, five with a deletion of R3-R4 region (also known as region C) and one with an insertion of one extra R3-R4 repeat. In addition, two deletions were detected, one of 1bp at position 27495, and another of 2 bp at positions 26372-26373 of the U29185 reference sequence.

A total of 12 of the 22 polymorphisms were singletons (nine SNPs + three indels), all of which were confirmed by two independent PCR amplifications and sequencing both strands. Among the 18 SNPs, transitions (12) were more frequent than transversions (6). Nine of the polymorphisms were found in the coding sequence (cds) and 13 in the rest of exon 2. In the well-characterised open reading frame (ORF) we found a number of previously described length polymorphisms and SNPs. These include the R2 deletion the region C deletion ,and SNPs in codons 117, 129, 142, 171, 177, 219 and 232. The codon 142 SNP was observed in four heterozygous individuals from Sub-Saharan Africa, while the codon 232 SNP was detected only in one Japanese sample.

We detected 13 previously unreported polymorphisms in the PRNP exon 2: 26215 (A→ T), 26281 (G→A), indel 26372-36373 (TA), 26611 (T→C), 26803 (T→C), 26883 (C→G), 26893 (A→G), 26908 (A→G), 27011 (C→T), 27079 (T→G), indel 27495 (T), 27623 (T→A) and 27802 (G→C).

The diversity and genealogy of PRNP haplotypes

The 22 polymorphisms define 28 different haplotypes (Fig.2). For the chimpanzee, only positions that are polymorphic in the human sequence are shown. Independently, we estimated allele frequencies with Arlequin 2.000, and both sets of frequencies were highly correlated (r=0.97 - 1).Haplotype diversity for the entire data set was relatively

low at 0.551. No less than 87.7% of the total haplotype variation can be accounted for by only two haplotypes, S1 and S2, which differ only at codon 129. Figure 2 reveals the existence of two major haplotype groups, one derivative of S1 (with 129Met) and the other of S2 (with 129Val). Haplotype diversity was 0.551, which is also a cosenquence of a haplotype frequency spectrum dominated by two haplotypes. Nucleotide diversity (π, the average heterozygosity across all polymorphisms in the sequence) was 0.00031. This value is very low compared to other genes that have been sequenced in a sufficient number of diverse individuals (Stephens et al. 2001). The $\theta\pi$ estimator was 0.74, which corresponds to the mean number of mutational differences between all pairs of sequences in the data set. The θw statistic, computed from the number of segregating sites (S) in a given locus and considering only SNPs, was estimated as 2.8.

Figure 3 shows a median-joining network, describing the most plausible mutational routes between the 28 haplotypes. Assuming that the phase of polymorphisms was correctly resolved (easy in this case as the number of variants is small and are well structured), it can be concluded that reticulations in the network represent either homoplasy (recurrent mutation events) or recombination events. The phylogenetic structure around haplotypes S1 and S2 is clearly star-like, with S1 having a slightly higher proportion of haplotypes derived by a single mutational step. There are only four single reticulations, involving positions 129- 219, 219-26372, 129-26908, and 26893-117. All are transitions and may therefore be attributable to recurrent mutation events, although crossing-over or gene conversion events cannot be ruled out. It is interesting to note that the susceptibility polymorphisms at codons 129 and 219 are involved in three out of four cases, indicating perhaps that one or both of them may have a recurrent origin. Positions 117, 129, 219 and 26893 are in CpGs.

RESULTS

Testing for the impact of natural selection

Tajima´s D statistic was -1.85 for the combined sample, excluding length polymorphisms, which is significantly smaller than expected under neutrality [which demographic scenario?] ($p<0.05$). When the length polymorphisms are included a D value of -2.02 is obtained ($p<0.05$). When statistical significance is evaluated by means of coalescent simulations with recombination ($R=2.79$), significance increases ($p<0.001$).

A positive value of D points to the possible impact of balancing selection or population subdivision on haplotype diversity, while a negative value can indicate either recent positive selection, a population expansion, or purifying selection on slightly deleterious alleles (Tajima 1989). Thus, there is a problem with this and other similar tests of conflating the signal of selection with the impact of demographic factors. In order to distinguish between these alternatives, we compared our results with values of Tajima's D obtained for 292 autosomal genes (Stephens et al. 2001). Figure 4 shows the resulting distribution, demonstrating that the value for *PRNP* falls in the left-hand tail of the distribution, with only 2.3% of the genes with lower values.

When compared with a data set of 132 sequenced genes (Akey et al. 2004), a similar D value is found for the EPHB6 gene, that was found to have a strong signal of positive selection in Europeans along with seven other genes from that data set. The value found for EPHB6 was clearly interpreted as the footprint of positive selection even after correcting the results for multiple tests.

Table 2 also presents results based on the D* and F* statistics from Fu and Li, which can be used to evaluate deviations from neutral expectations. A highly significant deviation from neutral expectations is obtained when length polymorphisms are included ($D^* = -3.77$ and $F^* = -3.69$ with $p<0.002$ and $p<0.001$ respectively). Less

significant results are obtained when length polymorphisms are omitted (D* =–3.65, $p<0.02$, F* = -3.54, $p<0.02$), but nonetheless they lead us to reject the neutral null-hypothesis.

Fu´s F test is negative for the combined worldwide sample (-47.28), and P values are highly significant, using coalescent simulations with and without recombination. Fay and Wu´s H for the combined sample is -1.37, a value that falls inside the null-distribution derived from coalscent simulations under the neutral model.

The same neutrality tests were applied independently to the haplotypes belonging to each of the two main clusters S1 (129Met) and S2 (129Val) shown in Figure 3. Table 2 shows the results of these analyses (in the columns labeled Met and Val), which were negative in all cases, and highly significant for the S1 haplotypes, but non-significant for the S2 haplotypes. Only one positive value was found, Fay and Wu´s H for the S1 cluster (non-significant), while in the S2 cluster it was negative and close to significance. Overall, then, the S1 cluster is indicative of a positive selective expansion, while the pattern of diversity in the S2 cluster is more consistent with the neutral expectation.

Table 3 shows the results of the McDonald and Kreitman (MK) test of neutrality applied to the combined worldwide sample using a range of different primate species as outgroups. While most results were non-significant, comparisons with the gibbon, siamang and spider monkey yielded significant results. However, we note that the MK test lacks power in this case, mainly due to the low level of polymorphism in the human PRNP gene, which is mainly non-synonymous.

Fay and Wu´s H test was applied for the combined sample, with non-significant results (Table 2). This showed that there is not an excess of high-frequency derived alleles relative to the ancestral state of alleles determined by comparison of humans and

the chimpanzee are within the range predicted by the neutral model and that no recent selective sweep can be detected. When this test is applied to S1 and S2 haplotype clusters non-significant values were also found.

The pairwise distribution of mutational differences

Figure 5 shows the mismatch distribution obtained from the combined worldwide sample as a smooth wave, such as that predicted for a population that has undergone a demographic expansion (Rogers and Harpending 1992) but with the mode at zero differences followed by a monotonic decrease, indicating that the expansion is recent in time. The distribution of mutational differences between all pairs of sequences was unimodal both for the total sample and for all geographic groups. We note that this is more consistent with a scenario of positive selection than one of balancing selection as proposed by (Mead et al. 2003).

Time to the Most Recent Common Ancestor (TMRCA) and Mutation Ages

We estimated the substitution rate in exon 2 as the number of differences divided by 2tL, where L is the sequence length and t the divergence time between species. We observe 26 differences between humans and the chimpanzee, yielding a divergence of 1.09%, a value that is similar to many others published in the recent literature (refs). Assuming a divergence time of 6 million years between humans and chimpanzees, the average substitution rate then becomes 0.91×10^{-9} per nucleotide per year for the PRNP exon 2.

By resolving the few network reticulations shown in Figure 3 (see methods), it was possible to use the Genetree software to estimate the age of the tree and of individual mutations using a coalescent approach. The tree shows a clear star-like shape

within the two main haplotype clusters (Figure 6). The estimation of TMRCA for the entire tree under a constant population model is 380 ± 105 thousand years (ky), which is in the low range of estimations for other human gene genealogies (Harris and Hey 2001); (Martinez-Arias et al. 2001) . The estimate obtained under a model of population growth was 200 ± 50 ky.

The polymorphism at codon 129 polymorphism is the oldest in the tree with an age estimated at ~200± 100ky under a constant population model and close to 100± 65kyunder population growth. Both dates are quite recent for the oldest polymorphism at a human autosomal locus.

Geographic variation of diversity and selection statistics

When examined by geographic region, the number of segregating sites and heterozygosity indicate that sub-Saharan Africans harbor the greatest diversity, while Native Americans exhibit by far the greatest homogeneity, with only the two basic S1 and S2 haplotypes that differ by the polymorphism at codon 129.

In most cases, the neutrality tests for individual geographic regions are non-significant (Table 2), a result that is most likely due to the lack of power caused by small sample size. It is interesting to note that under population substructure, Tajima's D is expected to be positive. However, for most neutrality tests (D, D*, F* and F) all continents yield negative values (reaching significance in a few cases) except for America, where they have non-significant positive values. Thus, substructure is not causing positive values and if substructure was affecting the obtained D values, the real ones would be even lower.

Moreover, Fay and Wu´s H for Americas is very close to significance (p=0.051).

RESULTS

Geographic heterogeneity

F_{ST} values calculated for the two SNPs in codons 129 and 219 were very different for the combined worldwide sample (the entire HGDP-CEPH panel): 0.16 for codon 129 and 0.01 for 219. Within continental regions the F_{ST} values for the same SNPs were, in all cases, lower than 0.1, with the maximum in Amerindian populations (0.09) for codon 129 and smallest values in Europeans and Central/South Asians (Table 4a).

An analysis of molecular variance (AMOVA) was performed to apportion the haplotype diversity among and within population groups using, on the one hand, the two codon SNPs (129 and 219) and, on the other hand, the set of full sequence data (Table 4a). Differences among large groups account for 15.32% of the total variation, which is more than the average (10%) found in an extensive study of neutral markers (Barbujani et al. 1997) and even more than a STR survey with the same samples (3 to 5%; (Rosenberg et al. 2002). Differences among populations within regions account for a mere 2.14% of the variation. The region showing the greatest differences in haplotype frequencies between populations is America, with highly significant P value, and this is reflected by the high Fst value found for polymorphism 129 in this group.

Similar analyses were performed for haplotypes generated from the sequence data (Table 4b). Fst values were low, in agreement with the low frequency in which most of the SNPs are found and AMOVA revealed relatively small differences among groups (8.25) and almost no differentiation among populations within groups (less than 1%). In America, there is no variation besides the 129 codon and the 16 chromosomes come from a single population.

The significance of the F_{ST} values found was empirically tested by comparing it to two distributions obtained with large numbers of genes: i), Akey et al. (2002) provided F_{ST} values for thousands of genes among three populations, and the value

obtained for *PRNP* was not particularly extreme; ii) Kidd et al. (2004) compiled allele frequencies and F_{ST} values for 369 SNPs (some scored as RFLPs) in a diverse population set that is comparable to the one we used. Thus, this is a more appropriate reference set, and against it the codon 129 F_{ST} value (0.16) is lower than that found in 28% of the polymorphisms. This means that the geographical variation of this SNP is not unexpectedly large.

DISCUSSION

In an effort to better understand the pattern and worldwide geographic structure of variation at the *PRNP* locus and the extent to which it has been shaped by natural selection, we have sampled and analysed the variation present in 348 chromosomes for exon 2 of this gene. By using a sequence-based approach, we have identified new SNPs, most of them in a very low frequency, indicating that this region is much conserved.

Selective forces acting on *PRNP*

Unraveling the role of selection in the past of the human genome is not a simple endeavor. The footprint of selection on a particular genomic region depends on a number of different factors, including the strength of selection, the type of selection (positive, balancing, purifying or some combination of these) and the time-scale within which selection has taken place.

The MK test was applied to evaluate the selective forces acting on PRNP from a phylogenetic point of view. While the results did not provide strong evidence for the action of selection, this may be largely attributable to the lack of power for the χ^2 statistic because of the limited amount of polymorphism. In a study by (Krakauer et al. 1998), rates of molecular evolution in prion gene were examined by means of the dn/ds ratio, with no evidence for the action of positive, directional selection on PRNP.

The Cercopithecidae, Hominidae and Cebidae cluster close to or below the line along which dn = ds. In primates, dn (0.037) / ds (0.137) is below one (0.27) indicating lack of positive selection in this branch. As no extreme value of this ratio was found, purifying selection, even being very evident, can not be considered to be a strong force driving the evolution in this interespecific analysis. This dn/ds ratio is very different from the one

obtained in *Bos taurus*, where purifying selection was found to be the main force driving the evolution of *PRNP* gene (Seabury et al. 2004) The time depth in which selection is acting can let different footprints in different species or populations. It is important to remember that evolution acting on primate branch can be different than the evolution acting in human populations. In this way, between-species comparisons like MK tests or dn/ds, can add information from an interespecific view, but add limited information when evolution in human populations is examined.

(Mead et al. 2003) proposed that balancing selection was an important force shaping the patterns of diversity at the *PRNP* gene both locally and globally in the human species. Our analyses have revealed no evidence to support this hypothesis. The pattern of variation in the PRNP exon 2, ascertained by sequencing, reveals frequency spectra of mutations and haplotypes that are not consistent with neutral expectations. However, the excess of low frequency variants (yielding, for example, large negative values of Tajima´s D) is indicative of either positive selection or purifying selection, but not balancing selection. Whereas the low number of non-synonymous variants indicates purifying selection, a recent study has shown that this form of selection does not yield large negative values of Tajima´s D (Gordo et al. 2002). Values like that obtained for *PRNP* (–2.02) were found only when simulations taking into account positive selection were applied. Other selection test statistics (D*, F*, F, H) lead to similar conclusions for these data.

As different local selection pressures can have influenced variation at this gene, analyses were performed for individual geographical regions. In general, the pattern of variation is similar to that observed for the combined worldwide sampke. Interestingly,

however, the neutrality test results for the Americas differ from those obtained for other regions. The Fay and Wu´s H tests showed that maybe in this region a relatively recent selective sweep can explain the increase in Val allele and the low diversity found in this region. Additional studies in Americas are needed to confirmed this hypothesis. This is also in agreement with the results found when Fay and Wu´s H test is applied in Val haplotypes. However, general Fst values were low and Fst values for codon 129 showed higher differences among human populations, but when compared with a Fst distribution obtained by (Kidd et al. 2004), it fall in a non extreme position.

In conclusion we can say that frequency-distribution tests provided significant statistical support for an excess of rare alleles and/or singletons in our overall sample for human *PRNP* exon 2 alleles. Significantly negative values for Tajima´s D and Fu and Li´s tests (D* and F*) are often interpreted as purifying or positive selection, but positive selection seems to be more likely because of the importance of the 129 polymorphism in prion diseases (that is unique to human species) and also because theoretical studies seem to confirm that these negative values of these statistics are obtained mainly with directional selection.

Perhaps in the Americas, where cannibalism has been better documented than anywhere else (confirmed in Anasazi (Marlar et al. 2000) and suggested in other groups (Gibbons 1997) (Defleur et al. 1999), the results can be explained by relatively recent selective sweep, favouring the increase of Val allele, reaching the highest frequency found in any other population. These could be similar as what was seen in Fore population, where Val allele seem to have a higher frequency than other populations close to the same geographical region.

These results are not in disagreement with a dynamic balancing selection, as different episodes of balancing selection in local regions during human history could arise with the same pattern of polymorphism that we observe today in human populations.
Two main conclusions arise from our analysis. First, the existence of an ancient, estable, balanced polymorphism can be discarded. Second, our results are consistent with a complex history of episodic or fluctuating selection; most likely short periods of positive selection with the associated selective sweeps, followed by drift and/or purifying selection. It is also possible that very short periods of balancing selection produce molecular signatures similar to a partial selective sweep (Navarro et al. 2000); (Navarro and Barton 2003), so such brief episodes are also consistent with our data.

Complex story, such as selective forces acting in different regions, and at an specific time in human history, seem to be a probable scenario for human *PRNP*.

Disease implications

Codon 142. This unconservative aminoacid change (Gly142Ser) may be involved in neurological disease causation, as it had been reported for the first time in a North African man with multiple sclerosis and in a Malian woman with viral meningoencephalitis (Laplanche, not published, see the Official Mad Cow Disease web page). In our sample, we have found this substitution only in Mbuti and Biaka Pygmies, Mandenka and San populations (one heterozygote individual in each population), for a total of 6% in Africans Similar frequencies for this change were reported by Mead et al. (2003), always limited to individuals of African descent. The relative high frequency of this 142 polymorphism advocates against any direct disease-causing role for this polymorphism.

Codon 232. This change (Met232Arg) in a position that lies in a region that is removed in the mature prion protein was first found as a compound heterozygote with a mutation on codon 180 mutation in a Japanese patient with prion disease by (Kitamoto et al. 1993), and a role in causing disease was suggested. However, later reports found this change in healthy individuals as old as 84 (Hitoshi et al. 1993; Hoque et al. 1996). We found it only in one out of 16 chromosomes from a sample of healthy Japanese individuals. Again, this appears to be a local, neutral polymorphism.

ELECTRONIC-DATABASE INFORMATION

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for PRNP [MIM: 176640], for CJD [MIM: 123400 and FFI [MIM: 600072]

Official Mad Cow-Disease Home Page, http://www.mad-cow.org

RESULTS

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Òscar Lao for statistical help. The raw data for the Fst comparisons were kindly provided by Mark Shriver (Pennsylvania State University) and Kenneth K. Kidd (Yale University.This work was supported by DGICT (BMC2001-0772) and DURSI (2001FI 00632 to M.S. and 2001 SGR 00285).

REFERENCES

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2:e286

Andrews NJ, Farrington CP, Ward HJ, Cousens SN, Smith PG, Molesworth AM, Knight RS, Ironside JW, Will RG (2003) Deaths from variant Creutzfeldt-Jakob disease in the UK. Lancet 361:751-752

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743-753

Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL (1997) An apportionment of human DNA diversity. Proc Natl Acad Sci U S A 94:4516-4519

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, et al. (2002) A human genome diversity cell line panel. Science 296:261-262

Cervenakova L, Goldfarb LG, Garruto R, Lee HS, Gajdusek DC, Brown P (1998) Phenotype-genotype studies in kuru: implications for new variant Creutzfeldt-Jakob disease. Proc Natl Acad Sci U S A 95:13239-13241

Collinge J, Sidle KC, Meads J, Ironside J, Hill AF (1996) Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. Nature 383:685-690

Defleur A, White T, Valensi P, Slimak L, Cregut-Bonnoure E (1999) Neanderthal cannibalism at Moula-Guercy, Ardeche, France. Science 286:128-131

Gibbons A (1997) Archaeologists rediscover cannibals. Science 277:635-637

Goldfarb LG, Petersen RB, Tabaton M, Brown P, LeBlanc AC, Montagna P, Cortelli P, Julien J, Vital C, Pendelbury WW, et al. (1992) Fatal familial insomnia and

familial Creutzfeldt-Jakob disease: disease phenotype determined by a DNA polymorphism. Science 258:806-808

Gordo I, Navarro A, Charlesworth B (2002) Muller's ratchet and the pattern of variation at a neutral locus. Genetics 161:835-848

Harris EE, Hey J (2001) Human populations show reduced DNA sequence variation at the factor IX locus. Curr Biol 11:774-778

Hitoshi S, Nagura H, Yamanouchi H, Kitamoto T (1993) Double mutations at codon 180 and codon 232 of the PRNP gene in an apparently sporadic case of Creutzfeldt-Jakob disease. J Neurol Sci 120:208-212

Hoque MZ, Kitamoto T, Furukawa H, Muramoto T, Tateishi J (1996) Mutation in the prion protein gene at codon 232 in Japanese patients with Creutzfeldt-Jakob disease: a clinicopathological, immunohistochemical and transmission study. Acta Neuropathol (Berl) 92:441-446

Kidd KK, Pakstis AJ, Speed WC, Kidd JR (2004) Understanding human DNA sequence variation. J Hered 95:406-420

Kitamoto T, Ohta M, Doh-ura K, Hitoshi S, Terao Y, Tateishi J (1993) Novel missense variants of prion protein in Creutzfeldt-Jakob disease or Gerstmann-Straussler syndrome. Biochem Biophys Res Commun 191:709-714

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241-247

Krakauer DC, Zanotto PM, Pagel M (1998) Prion's progress: patterns and rates of molecular evolution in relation to spongiform disease. J Mol Evol 47:133-145

Kreitman M, Di Rienzo A (2004) Balancing claims for balancing selection. Trends Genet 20:300-304

Lee IY, Westaway D, Smit AF, Wang K, Seto J, Chen L, Acharya C, Ankener M, Baskin D, Cooper C, Yao H, Prusiner SB, Hood LE (1998) Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res 8:1022-1037

Marlar RA, Leonard BL, Billman BR, Lambert PM, Marlar JE (2000) Biochemical evidence of cannibalism at a prehistoric Puebloan site in southwestern Colorado. Nature 407:74-78

Martinez-Arias R, Calafell F, Mateu E, Comas D, Andres A, Bertranpetit J (2001) Sequence variability of a human pseudogene. Genome Res 11:1071-1085

Mead S, Stumpf MP, Whitfield J, Beck JA, Poulter M, Campbell T, Uphill JB, Goldstein D, Alpers M, Fisher EM, Collinge J (2003) Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. Science 300:640-643

Navarro A, Barbadilla A, Ruiz A (2000) Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. Genetics 155:685-698

Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. Science 300:321-324

Palmer MS, Dryden AJ, Hughes JT, Collinge J (1991) Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease. Nature 352:340-342

Parchi P, Giese A, Capellari S, Brown P, Schulz-Schaeffer W, Windl O, Zerr I, Budka H, Kopp N, Piccardo P, Poser S, Rojiani A, Streichemberger N, Julien J, Vital C, Ghetti B, Gambetti P, Kretzschmar H (1999) Classification of sporadic

Creutzfeldt-Jakob disease based on molecular and phenotypic analysis of 300 subjects. Ann Neurol 46:224-233

Prusiner SB (1991) Molecular biology of prion diseases. Science 252:1515-1522

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552-569

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381-2385

Schneider, S., Roessli, D. and Excoffier, L. (2000), Arlequin ver. 2.0: A software for population genetic data analysis, 2.0 edn. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Seabury CM, Honeycutt RL, Rooney AP, Halbert ND, Derr JN (2004) Prion protein gene (PRNP) variants and evidence for strong purifying selection in functionally important regions of bovine exon 3. Proc Natl Acad Sci U S A 101:15142-15147

Shibuya S, Higuchi J, Shin RW, Tateishi J, Kitamoto T (1998) Codon 219 Lys allele of PRNP is not found in sporadic Creutzfeldt-Jakob disease. Ann Neurol 43:826-828

Soldevila M, Calafell F, Andres AM, Yague J, Helgason A, Stefansson K, Bertranpetit J (2003) Prion susceptibility and protective alleles exhibit marked geographic differences. Hum Mutat 22:104-105

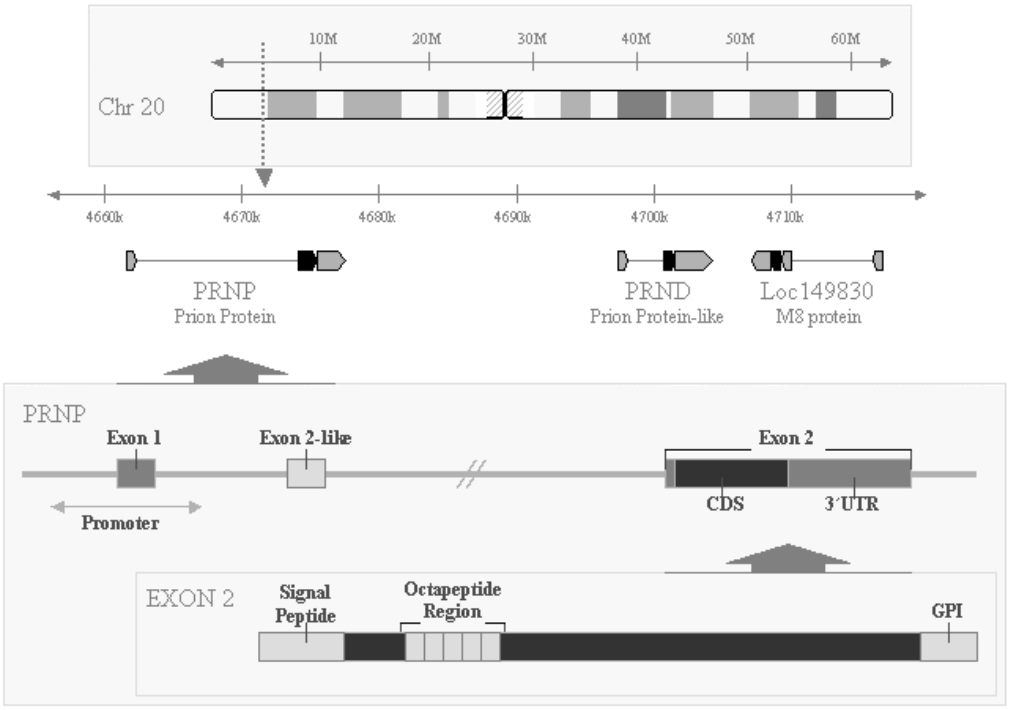
Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293:489-493

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

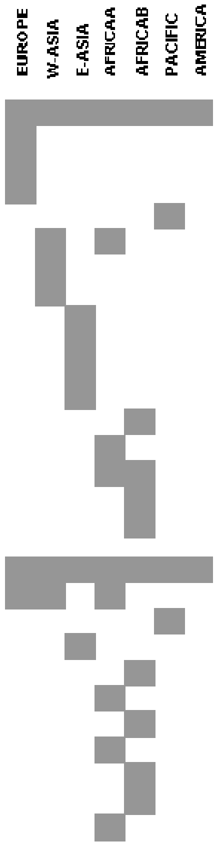
Tavare S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. Genetics 145:505-518

Valleron AJ, Boelle PY, Will R, Cesbron JY (2001) Estimation of epidemic size and incubation time based on age characteristics of vCJD in the United Kingdom. Science 294:1726-1728

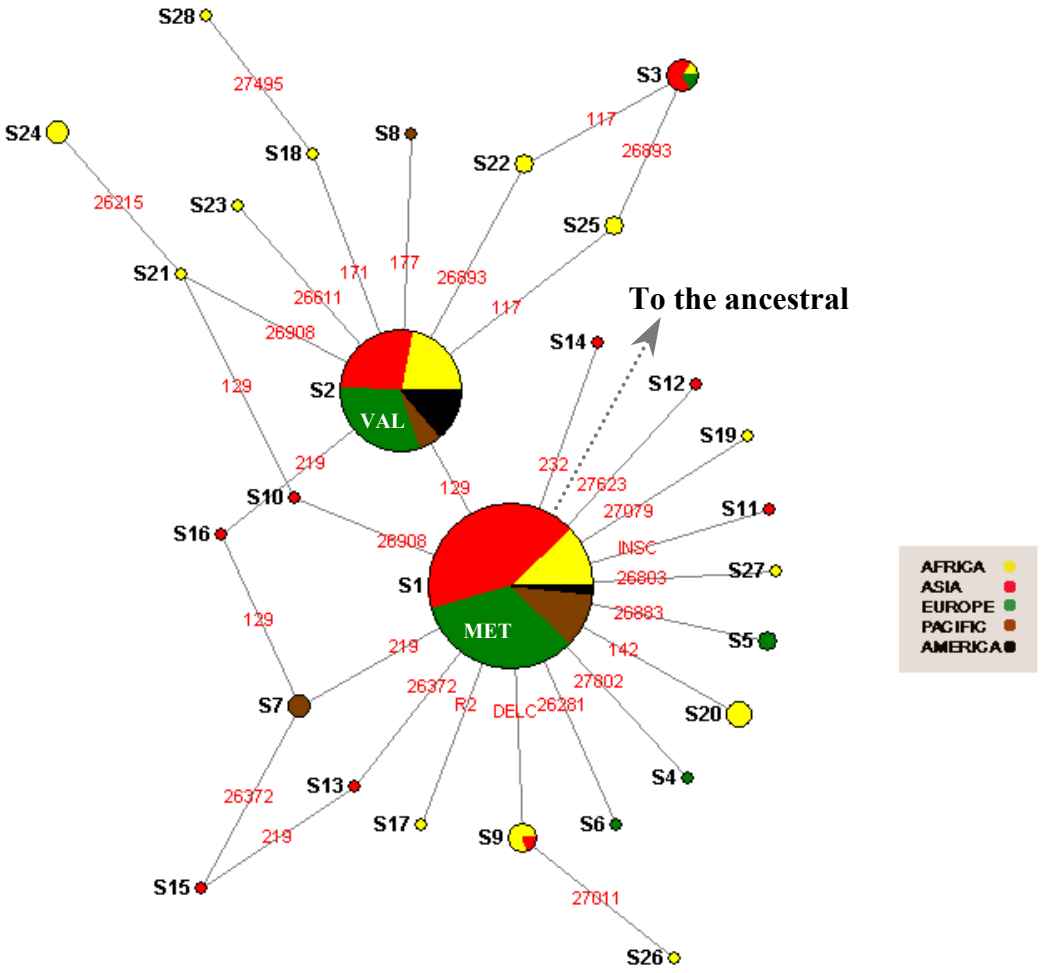
RESULTS

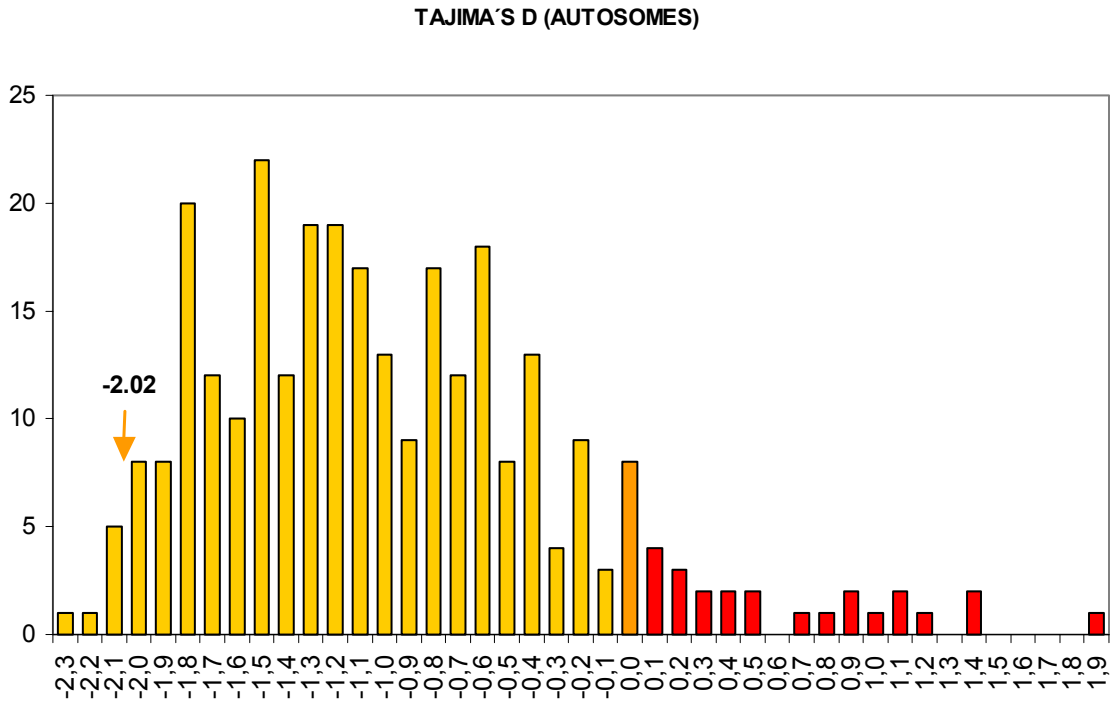

10M
20M
30M
40M
50M
60M
Chr 20
4660k
4670k
4680k
4690k
4700k
4710k
PRNP
Prion Protein
PRND
Prion Protein-like
Loc149830
M8 protein
PRNP
Exon 1
Exon 2-like
Exon 2
CDS
3' UTR
Promoter
EXON 2
Signal Peptide
Octapeptide Region
GPI

	R2	DELC	INSC	117	129	142	171	177	119	232	26215	26281	26372	26611	26803	26883	26893	26908	27011	27079	27495	27623	27802	
S1	R2	DC	IC	A	A	G	A	C	G	T	A	G	TA	T	T	C	A	A	C	T	T	T	G	217
S4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	1
S5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	2
S6	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	1
S7	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3
S9	.	–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5
S10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	1
S11	.	.	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
S12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	1
S13	.	.	.	.	.	.	.	.	.	.	.	.	–	.	.	.	.	.	.	.	.	.	.	1
S14	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	1
S15	.	.	.	.	.	.	.	.	A	.	.	.	–	.	.	.	.	.	.	.	.	.	.	1
S17	–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
S19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	1
S20	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4
S26	.	–	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	1
S27	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	1
S2	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	86
S3	.	.	.	G	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	6
S8	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
S16	.	.	.	.	G	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
S18	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
S21	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	1
S22	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	2
S23	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	1
S24	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	G	.	.	.	.	.	3
S25	.	.	.	G	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
S28	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	–	.	.	1
																								348
CHIMP	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	

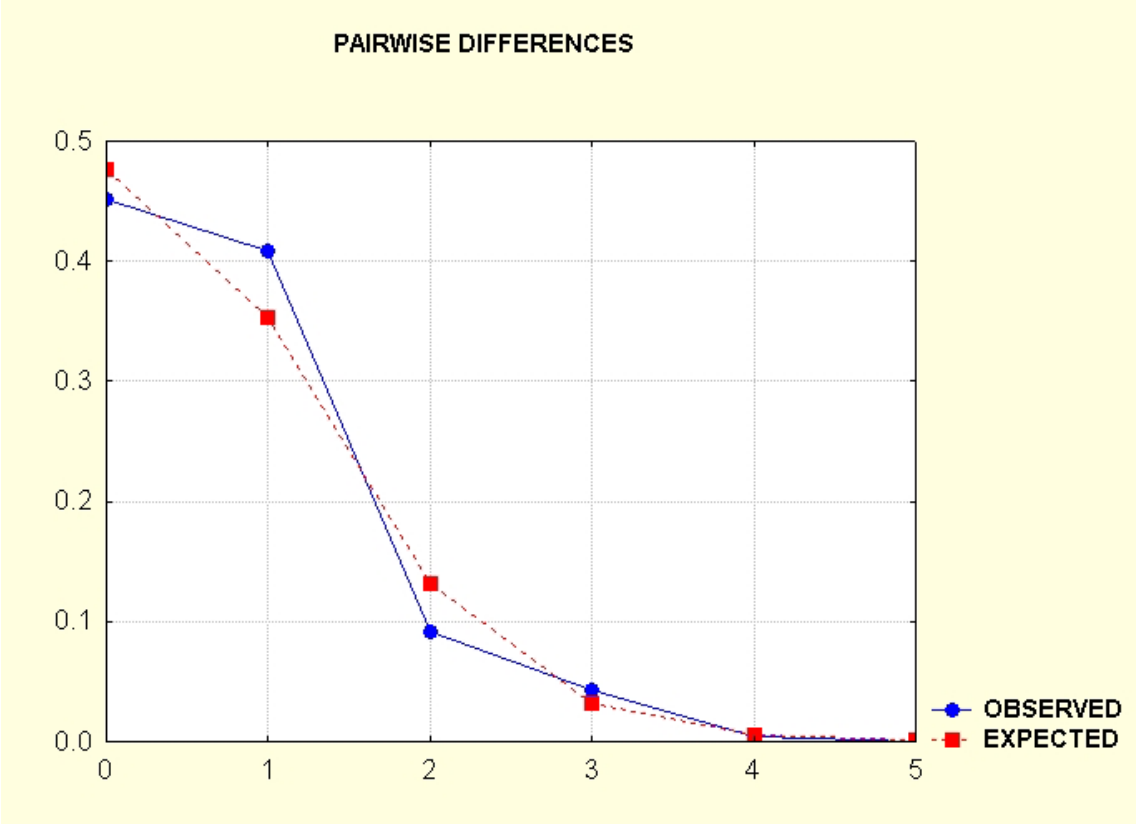


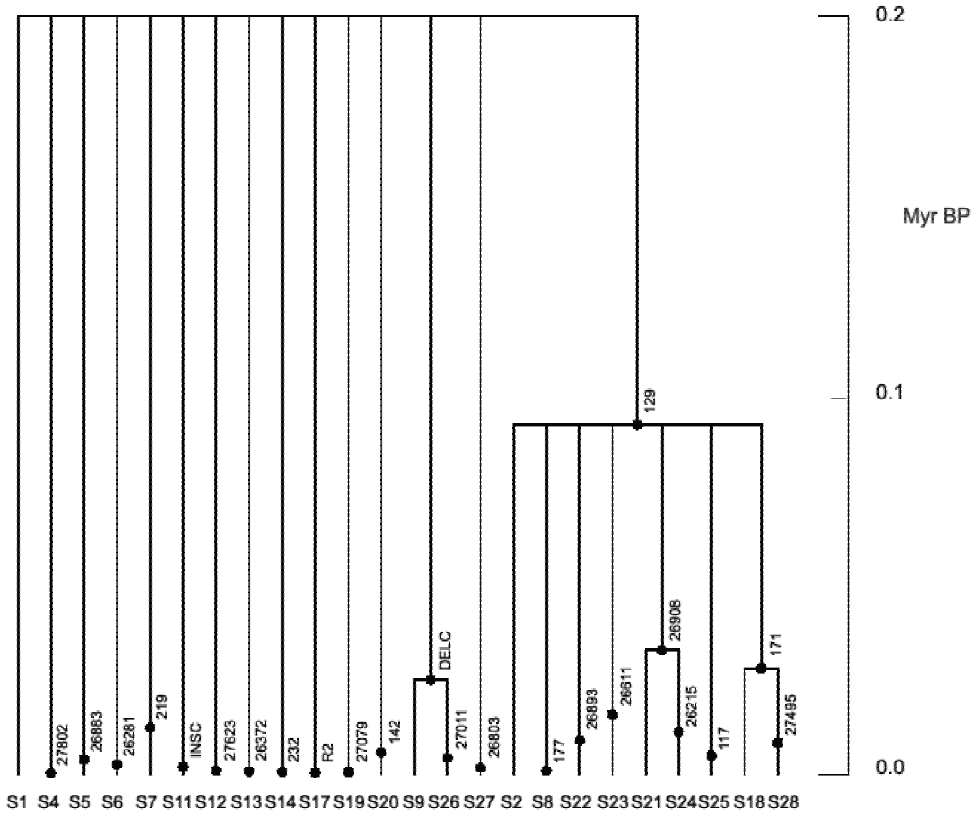
RESULTS


S28
27495
S3
S24
S18
S8
117
S22
26893
26215
S23
S25
S21
171
177
26893
26611
117
To the ancestral
26908
S14
S2
S12
129
VAL
S19
219
232
27623
S10
129
27079
S11
S16
26908
INSC
S27
S1
26883
129
MET
S5
219
142
26372
27802
S7
R2
DELC
26281
S20
S4
S13
26372
S17
S6
219
S9
S15
27011
S26
AFRICA
ASIA
EUROPE
PACIFIC
AMERICA

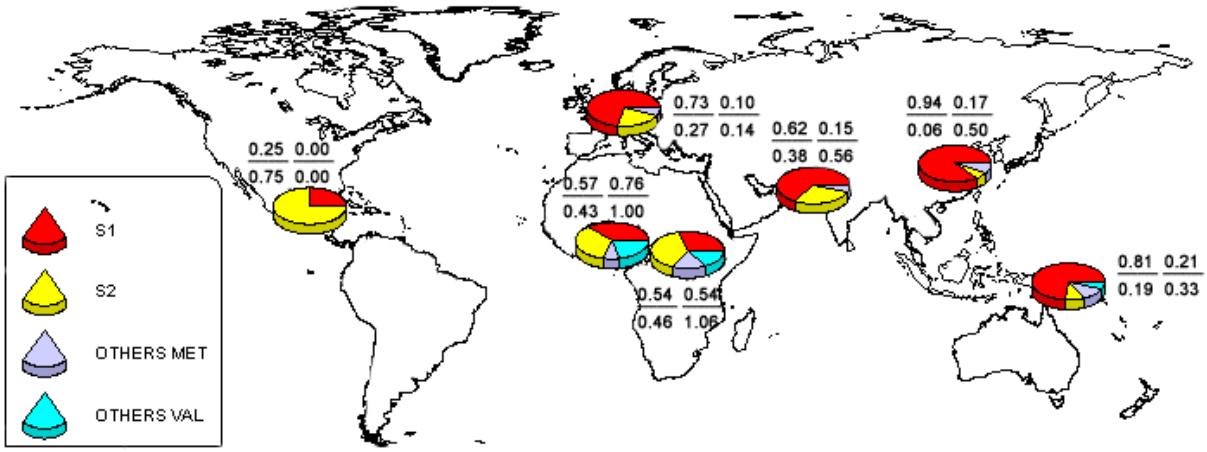

TAJIMA´S D (AUTOSOMES)
25
20
15
10
5
0
-2.02
-2,3
-2,2
-2,1
-2,0
-1,9
-1,8
-1,7
-1,6
-1,5
-1,4
-1,3
-1,2
-1,1
-1,0
-0,9
-0,8
-0,7
-0,6
-0,5
-0,4
-0,3
-0,2
-0,1
0,0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1,0
1,1
1,2
1,3
1,4
1,5
1,6
1,7
1,8
1,9

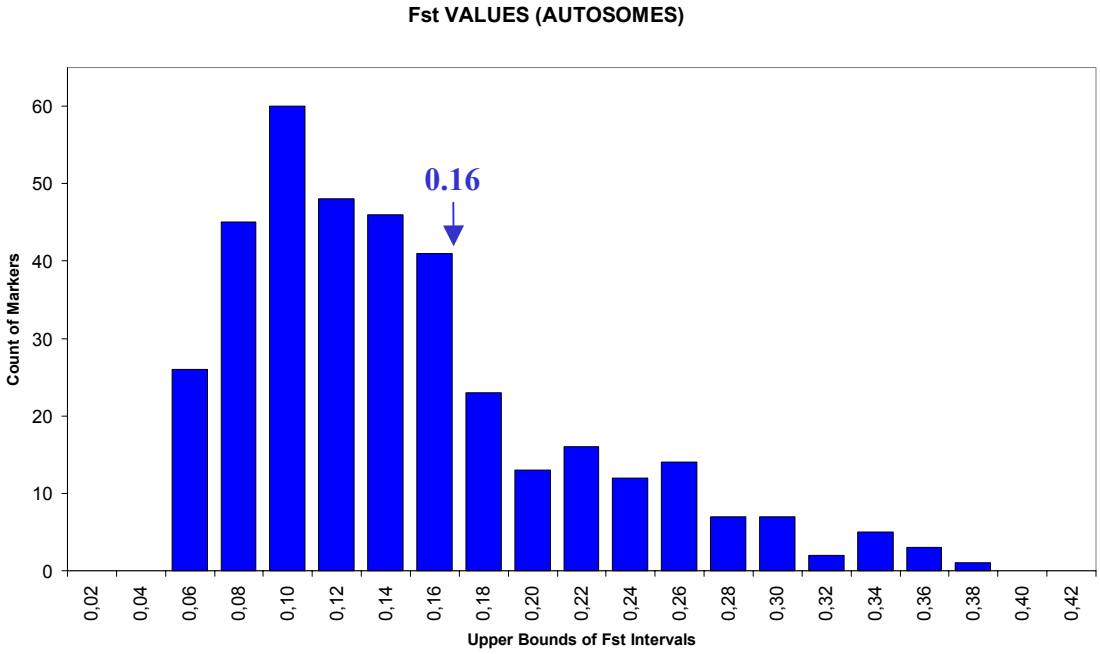
RESULTS


PAIRWISE DIFFERENCES
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
OBSERVED
EXPECTED


0.2
Myr BP
0.1
0.0
129
DELC
171
26906
27802
26863
26281
219
INSC
27623
26372
232
R2
27079
142
27011
26803
177
26883
26611
26215
117
27495
S1 S4 S5 S6 S7 S11 S12 S13 S14 S17 S19 S20 S9 S26 S27 S2 S8 S22 S23 S21 S24 S25 S18 S28

RESULTS


0.25 0.00
0.75 0.00
0.73 0.10
0.27 0.14
0.62 0.15
0.38 0.56
0.94 0.17
0.06 0.50
0.57 0.76
0.43 1.00
0.54 0.54
0.46 1.08
0.81 0.21
0.19 0.33
S1
S2
OTHERS MET
OTHERS VAL


Fst VALUES (AUTOSOMES)
Count of Markers
0
10
20
30
40
50
60
0.16
0,02
0,04
0,06
0,08
0,10
0,12
0,14
0,16
0,18
0,20
0,22
0,24
0,26
0,28
0,30
0,32
0,34
0,36
0,38
0,40
0,42
Upper Bounds of Fst Intervals

RESULTS

Figure 1.

Human PRNP gene. *CDS, UTR* and *GPI* stand for coding sequence, untranslated region and glycosylphophatidylinositol anchor, respectively.

Figure 2.

Haplotypes found in exon 2 of the *PRNP* gene. The first group of haplotypes contain Methionine (A) in position 129, while the second group contain Valine (G). The arbitrary names of the haplotypes are indicated on the left and the continental region where the haplotype is found on the right. The total haplotypic frequency is also shown. The alleles in chimpanzee are shown for the human PRNP variable sites.

Figure 3.

Median joining network of *PRNP* haplotypes. Circle areas are proportional to the frequency of the haplotypes. Branch lengths are proportional to the number of allele differences they represent. Haplotype number is shown next to the circles, and mutation positions are indicated on the branches linking two haplotypes. Each color inside the circle indicate the presence of the haplotype in an specific continental region.

Figure 4.

Tajima's D distribution from Stephen et al. 2001 based on 292 autosomal genes. The value obtained for PRNP gene is shown (-2.02).

Figure 5.

Pairwise distribution of all the samples. The results show no difference between observed (circles) and expected (squares) distribution.

Figure 6.

Coalescent tree of the *PRNP* haplotypes. Numbers along branches represent mutational events. Population growth has been taken into account. Tree was computed using θ_{ML}= 7.1, growth parameter = 3.8, N=10000 and generation time of 25.

Figure 7.

Haplotype frequencies in each continental group are shown. *Others Met* category includes all the haplotypes containing Met allele in position 129, except S1, that has its own category. *Others Val* category includes all the haplotypes containing Val allele in position 129, except S2, that has its own category.

Numbers in each continental region indicate the ratio of the frequency of Met allele *versus* the frequency of Val allele and next to it, the ratio of the MPD (Mean Pairwise Distribution) of Met haplotypes *versus* the MPD of Val haplotypes is also shown. (Met/Val MPDMet/ MPDVal).

Figure 8.

F_{ST} distribution from Kidd et al. 2004 based on 393 autosomal biallelic markers. The value obtained for codon 129 polymorphism is shown (0.16).

Table 1. Genotype and allele frequency distribution of codon 129 and codon 219 (a)

GROUP	2N	Samples Failed	129Met/Met	129Val/Val	129Met/Val	129Met	129Val
Africa	254	1	36.5	21.4	42.1	57.5	42.5
ME-NAfrica	356	2	50.6	11.4	38	69.6	30.4
Europe	322	1	46.9	7.5	45.6	69.7	30.3
SC-Asia	400	1	49.3	5.5	45.2	71.9	28.1
E-Asia	502	1	90.8	0.4	8.8	95.2	4.8
Pacific	78	0	61.5	10.3	28.2	75.6	24.4
America	216	2	15.1	45.3	39.6	34.9	65.1
GROUP	**2N**	**Samples Failed**	**219Glu/Glu**	**219Lys/Lys**	**219Glu/Lys**	**219Glu**	**219Lys**
Africa	254	0	100	0	0	100	0
ME-NAfrica	356	2	98.9	0	1.1	99.4	0.6
Europe	322	7	100	0	0	100	0
SC-Asia	400	2	97.5	1	1.5	98.2	1.8
E-Asia	502	1	94.4	0	5.6	97.2	2.8
Pacific	78	2	89.2	0	10.8	94.6	5.4
America	216	2	100	0	0	100	0

(a) These data were obtained by genotyping the whole HGDP-CEPH panel. The total number of chromosomes analysed was 2128 and the total of failures was eight for codon 129 and 16 for codon 219.

	Table 2. Neutrality tests for each continental region, all Methionine haplotypes (Met), all Valine haplotypes (Val), and for all the samples together (world) (a)											
	AFRICA	**AFRICAA**	**AFRICAB**	**EUROPE**	**W-ASIA**	**E-ASIA**	**PACIFIC**	**AMERICA**	**MET**	**VAL**	**WORLD**	**WORLD (COA,R=2.79)**
S	14	12	11	6	6	5	3	1	15	9	23	23
2N	68	28	40	104	64	64	32	16	243	105	348	348
Tajima's D	-1.34	-1.43	-1.32	-1.20	-0.87	-1.64	-0.60	0.65	-2.26	-1.72	-2.02	-2.02
Statistical significance	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S 0.10 > p > 0.05	N.S p > 0.10	N.S p > 0.10	Significant p < 0.01	N.S. 0.10 > p > 0.05	Significant p < 0.05	Significant p <0.001
Fu and Li's D*	-1.06	-1.29	-0.73	-2.73	-1.53	-0.93	-0.28	0.69	-4.72	-1.72	-3.77	-3.77
Statistical significance	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	Significant p < 0.05	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	Significant p < 0.02	N.S. p > 0.10	Significant p < 0.002	Significant p < 0.005
Fu and Li's F*	-1.38	-1.56	-1.08	-2.62	-1.55	-1.35	-0.43	0.77	-4.54	-2.03	-3.69	-3.69
Statistical significance	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	Significant p < 0.05	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	Significant p < 0.02	N.S. 0.10 > p > 0.05	Significant p < 0.001	Significant p <0.001
Fu´s F	-17.61	-6.71	-7.85	-3.50	-1.41	-6.60	-0.93	0.87	-29.63	-9.52	-47.28	-47.28
Statistical significance	Significant p <0.0005	Significant p <0.006	Significant p =0.003	N.S 0.10 > p > 0.05	N.S p > 0.10	Significant p <0.002	N.S p > 0.10	N.S p > 0.10	Significant p <0.001	Significant p < 0.01	Significant p <0.001	Significant p <0.001
Fay and Wu´s H	-0.80	-0.73	-0.92	0.38	0.29	0.49	0.46	-0.80	0.23	-1.44	-1.37	-1.37
Statistical significance	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p > 0.10	N.S p =0.051	N.S p > 0.10	N.S. 0.10 > p > 0.05	N.S. p > 0.10	N.S p > 0.10

(a) Significance of Tajima's D and Fu and Li tests is obtained directly. Significance of Fu's F and Fay and Wu´s H is calculated by coalescent simulations using Dnasp 4.00.

For the world sample, values of p for all neutrality tests are also calculated by using coalescent simulations coalescent simulations plus recombination (COA, R=2.79)

Table 3. McDonald and Kreitman test comparing human and different primates as outgroups		
MK TEST	**VALUES (a)**	**p**
Apes		
Chimpanzee	4/1-2/5	0.242 N.S.
Gorilla	1/1- 2/5	1.000 N.S.
Orangutan	11/5-2/5	0.169 N.S.
Gibbon	12/1-2/5	0.007**
Siamang	12/1-2/5	0.007**
Old-World monkeys		
Colobus	19/7-2/5	0.071 N.S.
Presbytis	16/8-2/5	0.099 N.S.
Baboon	20/9-2/5	0.084 N.S.
Rhesus macaque	20/9-2/5	0.084 N.S.
AGM (aethiops)	18/10-2/5	0.199 N.S.
New-World monkeys		
Capuchin	24/10-2/5	0.079 N.S.
Spider	24/7-2/5	0.023*
Marmoset	23/9-2/5	0.076 N.S.

(a) Values of Synonymous Fixed / Nonsynonymous Fixed - Synonymous Polymorphisms / Nonsynonymous Polymorphisms found in each comparison are shown.

Table 4a. Fst and AMOVA values for SNP 129 and 219 in the seven geographical regions (obtained by Real Time PCR for the whole HGDP-CEPH panel)

Fst	SUB-SAHARAN AFRICA	MIDDLE-EAST/ N-AFRICA	EUROPE	CENTRAL/ SOUTH ASIA	EAST ASIA	PACIFIC	AMERICA	GLOBAL[a]
Nº. of populations	6	4	7	8	6	2	5	**38**
2N	254	356	322	400	502	78	216	**2128**
129	0.02	0.03	-0.01	0.01	0.04	0.07	0.09	**0.16**
219	0.00	0.01	0.00	0.03	-0.01	0.06	0.00	**0.01**
AMOVA								
Among populations	2.11	3.06	-1.21	1.06	2.04	6.81	9.14	**2.14**
p	N.S.	<0.02	N.S.	N.S.	<0.01	N.S	<0.001	**<0.001**
Within populations	97.89	96.94	101.21	98.94	97.96	93.19	90.86	**82.54**
p								**<0.001**
Among groups								**15.32**
p								**<0.001**

Table 4b. AMOVA values for the haplotypes found in the seven geographical regions (obtained by sequencing the exon 2 of PRNP gene)

	AFRICA A	AFRICA B	EUROPE	W-ASIA	E-ASIA	PACIFIC	AMERICA	GLOBAL[a]
Nº. of populations	2	3	5	4	4	2	1	**21**
2N	28	40	104	64	64	32	16	**348**
Among populations	1.46	4.16	0	0	1.96	3.70	-	**0.82**
p	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	-	**N.S.**
Within populations	98.54	95.84	100	100	98.04	96.30	-	**90.93**
p								**<0.001**
Among groups								**8.25**

[a] Global comparison of the seven geographical groups for each SNP.

N.S. is non significant (>0.05).

5.3.3 CHAPTER V

Soldevila M., Andrés A.M., Yagüe J., Helgason A., Ramírez-Soriano A., Vallés M., Martínez-Díaz P.I., Stefansson K., Bertranpetit J.

Analysis of the promoter and complete exon 2 of the prion protein gene in sporadic and familial cases: *the genetic factor*.

In preparation

Analysis of the promoter and complete exon 2 of the prion protein gene in sporadic and familial cases: *the genetic factor.*

Soldevila M[1]., Andrés A.M[1]., Yagüe J.[2], Helgason A[3]., Ramírez-Soriano A[1]., Vallés M.[1], Martínez-Díaz P.I[1]., Stefansson K[3]., Bertranpetit J[1].

1. Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra

Address: Dr. Aiguader 80, 08003 Barcelona, Spain

Telephone number: (34) 93 542 28 39 Fax number: (34) 93 542 28 02

2. Servei d´Immunologia, Unitat de Biodiagnòstic de la Malaltia de Creutzfeldt-Jakob I d´altres malalties per Prions, Institut d´Investigació Biomèdica August Pi i Sunyer (IDIBAPS), Hospital Clínic, Universitat de Barcelona,

Address: Villarroel 170, 08036 Barcelona, Spain

Telephone number: (34) 93 932 275 414 Fax number (34) 932 275 783

3. deCode genetics

Address: Sturlugata 8, IS-101 Reykjavik, Iceland .

Telephone number: (354) 570 19 00 Fax number: (354) 570 19 03

Corresponding author: Jaume Bertranpetit

e-mail: jaume.bertranpetit@upf.edu

Abstract

Prion diseases are rare neurodegenerative disorders that include, among others, Creutzfeldt-Jakob disease (CJD). This disease is unique in having three etiological forms: inherited, sporadic and acquired. It is still an enigma why more than 80% of the cases of prion diseases are sporadic. Some patients results in familial cases after resequencing analysis, but still, the percentage of sporadic cases are very high. Further studies, show that other genetic factors (regulatory regions or upstream of exon 1) may be important in the susceptibility of prion sporadic cases. We analyse promoter region and the complete exon 2 of PRNP gene for a total of 119 sporadic and familial cases in order to see any genetic factors that can explain prion diseases. No particular footprint for sporadic cases can be seen in exon 2, where very few polymorphisms are found, indicating that the region is very conserved. Maybe some particular cases can be explained by mutations in this region, but this is not a general trait. Another region of interest is the promoter, where three common SNPs (-563,-101 and +310) are found. Carriers of these mutations in sporadic CJD cases, familial cases and control groups are compared. Interestingly, the first two polymorphisms are mainly present only in European populations, but polymorphism +310 is found world-wide. Allele C in the +385 polymorphism has not been detected in any of our samples. Also some other new polymorphisms have been found by sequencing the whole promoter region, and haplotypes observed for sporadic and familial patients are provided. Studies in this promoter region help us to understand the genetic factor that underlies these prion diseases.

Introduction

The transmissible spongiform encephalopathy (TSE) diseases are a group of rare, fatal, and transmissible neurodegenerative diseases that include all prion diseases. TSE in humans are generally classify as hereditary (or familial), acquired (or infectious) or sporadic.

Familial cases have been related to mutations in the coding part of prion protein gene. The two most frequent mutations are located at codons 200 (Glu/Lys) and 178 (Asp/Asn). The E200K mutation represents the most frequent cause of familial Creutzfeldt-Jakob disease (CJD), and the D178N point mutation has been linked to two separate disorders: fatal familial insomnia (FFI) and familial CJD. If the mutation is coupled to Methionine at position 129 then it is associated with FFI, and if it is coupled to Valine at position 129 is associated with CJD.

Acquired cases have been found to be related to environmental source of infection such as the use of contaminated instruments during neurosurgical procedures or a history of treatment with pituitary-derived hormones (both cases called iatrogenic CJD). Also it includes cases of infection related with Bovine Spongiform Encephalopathies (BSE), causing variant or new variant CJD and cases of endocannibalism in a Papua New Guinea ethnic group, Fore, causing Kuru.

Little is known about the reasons of why a person that lack any family history of prion diseases or any other environmental source of infection can develop the sporadic form of Creutzfeldt-Jakob disease (CJD).

RESULTS

a) It has been estimated that 10% of apparently sporadic TSE cases show the presence of a *PRNP* mutation (e.g. at position 178 or 200) after a detailed resequencing study of the coding part of the cases (The EUROCJD Group(2001). (See for example (Goldman et al. 2004).

b) Also, some "de novo" mutations explain the lack of family history and it can be explained by the presence of some CpG dinucleotides, such as mutation at codon 178 (causing TSE) ((Windl et al. 1996; Dagvadorj et al. 2002), that produces genetic instability, but this represents a low frequency of cases. As a general rule, the mutation rate for a CpG can be established in the order of five times the base mutation rate (Krawczak et al. 1998).

c) Homozigosity (for methionine or valine) at a common human polymorphism (codon 129) has been related with an increased susceptibility to developed sporadic prion diseases (Palmer et al. 1991).

d) Other genetic susceptibilities can come from the promoter region. McCormack (McCormack et al. 2002), found in the upstream regulatory region a polymorphism (-101G) that can be influencing the susceptibility to develop sCJD. Together with other polymorphisms (+310C and +385C) that were found in the intronic region, they shown that carriers of these mutations are more frequent in the sporadic group than in controls. They suggest that polymorphisms in the regulatory region of the PRNP gene may be a risk factor for CJD, but larger numbers of patients and controls were required to confirm this hypothesis.

e)It has also been found (Mead et al. 2003) a significant independent association between sCJD and a polymorphism upstream of *PRNP* exon 1 in addition to the strong susceptibility conferred by codon 129. However, no association was found between these polymorphisms and vCJD or iatrogenic CJD.

Traditionally only the coding part (cds) of exon 2 is sequenced to assess the familial or sporadic origin. As we have seen for the regulatory region, maybe other mutations in other parts of the gene, like promoter, exon 1 or the 3´UTR region of exon 2 could explain some sporadic cases.

One previous study has analysed this complete exon 2 region in patients and no differences with the control sample have been observed, except the known mutations in familial cases in the coding part (Mead et al. 2001). However, this study genotyped some polymorphisms found in the control population, maybe undetecting polymorphisms unique to patients suffering prion diseases.

Only very short number of polymorphisms have been detected in the 3´UTR exon 2 in world wide control samples (Mead et al. 2001; Soldevila et al. 2003), indicating that the level of conservation of this region is very high and that changes in this region could affect the stability of the mRNA or the normal function of the protein.

In this study, promoter and the whole exon 2 of prion protein gene (PRNP) have been resequenced in familial and sporadic cases in order to detect the patterns of nucleotide variation in both groups (Fig. 1). Moreover, the patterns observed can be compared with control sequences and assess the origin of some mutations associated with familial diseases.

The goal of the present study is to analyse the genetic variation in the prion protein gene by sequencing a long fragment of the gene encompassing the promoter region (which overlaps exon 1) and all exon 2 to ascertain the possible existence of susceptibility alleles or haplotypes and analyse this variation in an evolutionary framework.

Patients and Methods

Patients and diagnostic criteria

The samples were obtained by the Catalan surveillance system (see (Sanchez-Valle et al. 2004) for details on the patients). A total of 119 patients with signs of human prion diseases were included in this study (only familial and sporadic cases).

Diagnostic criteria used for human prion diseases was done following World Health Organization consultation criteria (1998). The WHO, in 1998, established for the diagnostic of the human prion diseases, the categories of definite, probable and possible. Neuropathology at present provides the only means of establishing a diagnosis of definite sporadic cases. In the inherited forms (or familial), the genetic analysis is the only way to confirm the diagnosis of inherited prion disease by the presence of a pathogenic mutation in the *PRNP* gene.

In the genetic cases, particular mutations are associated with CJD (Creutzfeldt-Jakob disease), FFI (Fatal familial insomnia) or GSS (Gerstmann-Straussler-Sheinker). As phenotype heterogeneity is found among specific mutations, classification of the patients was only based on the genetic evidence. Thus, patients with the mutation in position 178 coupled to 129Methionine polymorphism were classified as FFI, and if it was coupled to 129Valine were classified as fCJD. Mutations in position 200 of the PRNP gene are only associated with fCJD. Patients with pathogenic number of repeats in the octapeptide repeat region were classified as *others* inside the familial category. Any GSS patient is present in our sample.

The clinical data were registered as reported by the referring neurologists, following the questionnaires proposed by the European and Allied Countries collaborative study group of CJD (EUROCJD, 2004).

Informed consent was obtained for the genetic analysis and the study was approved by the Ethical Committee of Hospital Clínic and Universitat Pompeu Fabra, Barcelona.

The total of 119 patients were sequenced for the complete exon 2 and promoter (from position –600 to +139, 12034-12772).

A subset of 69 patients (24 familial cases and 45 sporadic) was also sequenced from +109 to +684 (12742-13316) of the promoter region, including the whole exon 1 of the *PRNP* that overlapps with the promoter.

For three patients we also had other family members that were also sequenced.

We also analysed the frequency distribution of -563, -101 and +310 and M129V polymorphisms in a control group (96 subjects from the normal population from Catalonia, and 43 Catalans and 43 Basques).

As a reference of variation, a total of 281 samples representing the variation of human populations (HGDP-CEPH panel) (Cann et al. 2002) were also analysed for the same region.

Sequencing

DNA was extracted from CJD patients whole blood with QIAamp_ DNA Blood Mini Kit Qiagen GmbH, according to the manufacturer procedure. If whole blood samples were not available, DNA was extracted from serum, CSF or paraffin-embedded tissue sections using the QIAamp ® DNAMini Kit Qiagen GmbH.

RESULTS

Complete PRNP exon 2 was amplified by PCR using the primers described in Soldevila et al. (Soldevila et al. unpublished results). The open reading frame was also amplified by PCR using primers described in Sanchez-Valle et al 2002, obtaining a duplicate for this region, and an excellent accuracy in the genetic diagnostic. The promoter region (12034-12772) –600 to +139 including exon 1 of the *PRNP* was amplified using primers described in (McCormack et al. 2002) (UP3F-UP3R) and also using new primers specially design for this study (DP1F-DP1R). Primers and conditions are available on request. Segments were sequenced by use of the BigDye Terminator Cycle Sequencing kit from PE Biosystems on an ABI 3700 (Applied Biosystems, Foster City, CA, USA) DNA sequencer. The same primers were used for cycle-sequencing reactions. 100% of the samples were sequenced for both strands, which provided an overlap between all the PCR segments. The sequences were aligned with SEQMAN II 4.03 (DNASTAR) and manually checked.

A subset of 69 patients were also sequenced for the promoter region from +109 to +684 (12742-13316) using INT1F-INT1R primer from (McCormack et al. 2002).

SNPs were characterized and were designated according to their position on clone U29185 (Lee et al. 1998). SNPs were also designated given specific names; for example, ORF polymorphism *PRNP* M129V, or an SNP that is 101 bp 5′ to *PRNP* exon 1, was designated "-101," and a SNP that is 310 bp 3′ to PRNP exon 1 was designated "+310". This is in agreement with other articles published analysing the same region (Mahal et al. 2001; McCormack et al. 2002).

Cloning

PCR products with extra or deletion of repeats in the octarepeat region were subjected to pMOSBlue blunt ended cloning kit (Amersham Biosciences) following the manufacturer's instructions. The fragments with indels were subjected to cloning in order to ascertain the length and composition of each insertion or deletion. Seven microlitres of PCR product were treated with pK enzyme mix, incubated at 22ºC for 40 min and ligated into pMOSBlue vector overnight. 1.5 ml of the ligation product were transformed into 20 ml of competent cells, grown in 80 ml of SOC medium at 37ºC during one hour and plated on IPTG/X-gal agar plates. After 16 hours, white colonies were subjected to direct PCR screening using the same primers used in the PCR amplification. Inserts were identified by agarose gel electrophoresis, purified and sequenced.

Some of the samples with extra or deletion of repeats in the octarepeat region were subjected to another protocol (see(Sanchez-Valle et al. 2004) in a different laboratory obtaining in some cases independent results to confirm the sequence.
In order to confirm if 129 position was coupled to mutation at position 178, a restriction assay was performed on the cloning products. Restriction enzymes Tth 111 and Tai I were used.

Data analysis

Haplotype inferences and median-joining networks

RESULTS

Haploid sequences for exon 2 were obtained directly taking into account the information obtained in the cloning method. In three cases some family genetic information helped in the phase assignment. Also, haplotypes were estimated using the bayesian algorithm as implemented in Phase version 2.0 package (Stephens and Donnelly 2003) and compared with the results obtained directly. Just a few differences were found and easily resolved when all the information was put together.
For the promoter region only Phase information was available.

Network 4.1.0.0. analysis software was used to establish median-joining networks among the haplotype of our sample set (Bandelt et al. 1995).

Statistical Analysis

Associations were analysed by using the X^2 test. When the expected frequency was below 1 or if the expected frequency was less than 5 in more than 20% of the cells, Fisher exact test was applied.

Results

From a total of 90 cases of sporadic CJD, 32 cases were classified as definite, 19 as probable and 38 as possible following the WHO consultation criteria (Table 1). One case was classified as others because no clear sporadic or familial assessment was possible, even if necropsy confirmed CJD diagnosis. This individual was homozygote for two extra insertions of R2, even if it can also be the result of a deletion in one chromosome, as no polymorphic position was found in the region studied. This patient was also used in another study (Sanchez-Valle et al. 2004). Four patients that were classified as sporadic showed non-pathogenic polymorphisms in the octapeptide repeats region.

A total of 29 familial cases were detected (Table 1). Ten patients carried the Lys allele in codon 200, all of them in heterozygosis, were diagnosticated as CJD. 18 patients carried the Asn allele in codon 178 in heterozygosis coupled to a Methionine in codon 129, and they were assigned to be FFI. Another familial case was obtained (CJDf-others), but in this case, disease was associated to a eight extra repeat insertion associated with Valine in codon 129.

PRNP Nucleotide Sequence Variation

For the total of 119 cases a sequence of 739 bp was obtained for the promoter region and 2378 bp for the exon 2.

A total of 11 SNPs (single nucleotide polymorphisms) have been found in the complete region of exon 2. Also, extra repeats and deletions in the octapeptide repeat region have been found in exon 2. The variants found are A117A, G124G, M129V, D178N and

RESULTS

E200K in the open reading frame and 26281, 26417, 26883, 26893, 27274 and 27471 in the 3´UTR of exon 2 (designated according to their position on clone U29185 (Lee et al. 1998). The last two variants have never been observed before. The others have been detected in world wide control sample studies (Soldevila et al, submitted) (A117A M129V, 26281, 26417, 26883 and 26893) or in cases studies (M129V, D178N and E200K). 26417 has also been observed in the Icelandic population (Soldevila et al., unpublished results).

In the promoter region two substitutions have been found in patients suffering prion diseases. These polymorphisms are located in the position 12071 (-563) and 12533 (-101). Position –563 has never been described before. These two polymorphisms have been also found in HGDP-CEPH samples, but mainly restricted to European populations (Fig. 2).

In the subset of 69 patients where positions from +109 to +684 (12742-13316) were also sequenced, polymorphism +310 (12942) has been found, whereas another position described by McCormack et al., +385 (13017) was found to be monomorphic.
Another polymorphic position, 13045 (A->C) carried C in all the samples used in this study. Position 13270 (A->G) carried G in all the samples, but in some cases it was not assessed because it is located at the end of the fragment amplified and some sequences were not long enough to check it properly. This last position is not considered.
Two new low frequency polymorphisms were found in this region. 12875 (G->T) and 12890 (G->A). Only 12875 was found to be polymorphic in HGDP-CEPH samples.

Polymorphism in codon 129

Genotypic and allelic frequencies for polymorphism in codon 129 (Met/Val) for each group of patients are shown (Table 2). Among 69 sCJD (definite and possible) patients, 50.7% were homozygous for Methionine (Met/Met), 29.0% were heterozygous Met/Val and 20.3% were homozygous for Valine (Val/Val). This results are similar to a previous study (Sanchez-Valle et al 2004) that were significantly different from control samples (Saiz et al. 2001).

Eight extra repeats in a familial CJD patient

An eight extra repeats in the octapeptide repeat region coupled to Valine in codon 129 was observed in a patient in 1998 in Valencia (Spain). The other allele had a normal number of repeats (5) and Methionine allele in codon 129. The structure of the eight extra repeats is R1-R2-R2-R3-(R2-R2-R2-R2-R2-R2-R2-R3)-R4 coupled to Valine in codon 129. This insertion was similar to the one described by (Laplanche et al. 1999) but aminoacid found in position 129 in this case was Methionine. This is the first time this pattern has been described. Other eight extra repeats have been described coupled to Valine, but the pattern of the repeats was different (Goldfarb et al. 1991; van Gool et al. 1995).

Exon 2 haplotype reconstruction and phylogenetic relationships

Exon 2 haplotypes were obtained for patients suffering prion diseases (Fig. 3). Haplotypes termed S were present in an study of worldwide samples (Soldevila et al, unpublished results). Haplotypes termed M are only present in sporadic patients and haplotypes referred as D are only present in familial patients.

RESULTS

The phylogenetic relationship between each haplotype is shown (Fig.4). Most of the haplotypes found in patients are derived from S1 (that is characterized by a Methionine in aminoacid 129 of *PRNP)*. All the chromosomes carrying mutation 200 (CJDf) are identical for the complete exon 2, and they constitute the haplotype D35.
All the chromosomes carrying mutation 178 are identical for the exon 2 region (haplotype D36) except one chromosome that constitutes haplotype D38, where a deletion of 24 base pairs is found in the octapeptide repeat region.

Differences in exon 2 between patients and controls

None of the SNPs or haplotypes found in exon 2 of the PRNP gene can explain the differences found between sporadic patients and controls. As we can see from the network (Fig. 4), there is not an specific mutation that distinguish the sporadic group from the others. Only some mutations are found exclusively in patients with sporadic CJD, but the amount of patients having these mutations are very limited. However, this mutations could be neutral, and further studies should be performed in general population to confirm these polymorphisms are non-pathogenic.
This is the case of mutation 26417 that has been observed in another study with more resequencing results (Soldevila et al., unpublished results).

Promoter haplotype reconstruction

Haplotypes found in the whole promoter region based on 69 patients are being described for the first time (Table 3). Five SNPs are found which generate six different haplotypes for this region. P1 haplotype is by distance the most common haplotype.

Association with PRNP promoter polymorphisms with CJD.

In order to determine if regulatory region of PRNP is affecting the susceptibility to develop sCJD, linkage disequilibrium analysis with polymorphism in codon 129 (Methionine/Valine) was initially tested. To eliminate the effect of linkage disequilibrium with codon 129, and to better compare the groups with different 129 genotype frequencies, groups were subdivided by codon 129 genotype.
Among Methionine homozygotes (codon 129) (Table 4), some statistically significant comparisions are found between sCJD, fCJD and Europe groups. The comparison with control group is pending.

Control group results for polymorphisms in promoter region are still pending.
Genotype and allelic frequencies for polymorphism -563, showed no differences between groups in Table 4. Also carriers and non carriers for this variation, indicate that the differences between groups are not statistically significant. Thus, this SNP alone does not contribute to susceptibility to sporadic prion diseases.
Polymorphism at position –101, show statistically significant differences when sCJD and fCJD are compared with Europe group independently. Allelic frequencies, genotype frequencies and carriers and non carriers of –101 mutation, showed $p<0.02$. However, comparisons between sCJD and fCJD groups for -101 polymorphism showed no evidence of differentiation.
When all the groups are compared independently with the control group from (McCormack et al. 2002) (United Kingdom), carriers and non-carriers of –101 polymorphism, non-significant differences are found (not shown).

An excess of carriers of any of these mutations (or both) could be more susceptible to prion diseases, as statistically differences exist between all the groups.

RESULTS

In the 129 Methionine/Methionine subgroup there was a statistically significant excess of sCJD patients carrying a rare allele at –563 and/or –101 when compared to fCJD, where the number of patients with any of these mutations were less than in sCJD group. Significant differences were also found between sCJD patients and Europe group and when fCJD and Europe group are compared. Europe group showed a higher number of carriers versus non-carriers, whereas the other two groups showed a higher number of non-carriers versus carriers, specially in fCJD group.

A test was therefore performed where the numbers of individuals who carried any rare alleles were compared with those who were homozygous for the common allele at all two positions (Table 4). This showed that among 129 methionine homozygotes 7 of 27 sCJD patients but only 2 of 20 fCJD carried a rare allele (P=0.159). This result should be tested with controls of the same population (pending). The increased risk associated with carrying a rare allele (-563 or -101) may be due to a more complex model where two or more alleles independently increase susceptibility to sCJD (McCormack et al. 2002).

Discussion

Exon 2

Very few differences have been seen between sporadic samples and the other groups. Only some variants with a very low frequency have been found only in sporadic group. These mutations could represent low frequency variants with no further disease association. However, we can not discard the possibility that very few cases of sporadic diseases could be explained by mutations in this region and then they would be classified as familial cases.

In conclusion, we can say that polymorphisms in the complete exon 2 of PRNP gene could not explain most/all sporadic cases from a genetic approach, as no specific polymorphisms have been detected in this region of the gene. However, it is known that homozigosity (for methionine or valine) at a common human polymorphism (codon 129) has been related with an increased susceptibility to developed sporadic prion diseases.

Familial cases with Lysine in codon 200 and 178 were coupled to Methionine at codon 129. It has been seen that in some cases, the 200 Lysine has been coupled to Valine allele in codon 129 (Lee et al. 1999), but this is restricted to Western European Haplotypes (e.g. German, Sicilian or Austrian populations). The samples analysed are likely to correspond to the Mediterranean haplotype described by Lee et al. based on different microsatellites. All the chromosomes found with this mutation were identical for the complete exon 2, and they constitute the haplotype D35.

For polymorphism D178N, it has been seen that it is likely that there is a cluster of this mutation in the Basque Country (Zarranz et al. 2005). The samples used in this study

come from this region and we have seen that all the chromosomes carrying this mutation are identical for the exon 2 region (haplotype D36) except one chromosome that constitutes haplotype D38, where a deletion of 24 base pairs is found in the octapeptide repeat region coupled to Methionine at position 129. This finding suggests that among Basque haplotypes that include 178N mutation, short variation exist (Rodriguez-Martinez et al. 2005), but small differences can be found. All the chromosomes with 178N polymorphism are coupled to Methionine in codon 129, which from a genetic point of view has been related with FFI (Goldfarb et al. 1992), but the clinical phenotype is heterogenous (FFI and CJD).

Promoter

A detailed resequencing project provided us interesting data about the polymorphism in this region. Haplotypes found in patients are generated for the first time. Little variation is found in this region, according with limited variation seen in exon 2. The polymorphisms found in this region are mainly restricted to Europe. If having a high frequency of rare alleles of these SNPs (-563, -101 and +310) are related to sporadic CJD (pending), then this could be specially relevant for European populations.

Conclusion

Sporadic CJD disease is still not well understood. In this study, we wanted to provide more genetic data to be able to understand if any genetic factor underlined these sporadic cases. Only polymorphism in codon 129 show significance evidence to think that homozygotes for this SNP have an increase risk to develop sporadic prion diseases.

No specific SNPs were found to be relevant in exon 2 for the understanding of these pathologies. However, in promoter, polymorphisms -563, -101 and +310 seem to be a key factor, together with polymorphism in codon 129,

This means that not only, polymorphisms in the coding region can be linked to an increase susceptibility to sporadic prion diseases by alterations in the protein structure, but also polymorphisms in the promoter region can modulate the expression of the prion protein and thus, increase also the susceptibility to this pathology.

Acknowledgements

The authors thank all physicians and neuropathologists in Catalonia notifying suspect cases to the Catalan surveillance system, and providing data and samples. Special thanks to Raquel Sánchez-Valle for her help. This work was supported by DGICT (BMC2001-0772) and DURSI (2001FI 00632 to M.S. and 2001 SGR 00285).

.

References

(1998) Human transmissible spongiform encephalopathies. Wkly Epidemiol Rec 73:361-365

(2001) Genetic epidemiology of Creutzfeldt-Jakob disease in Europe. Rev Neurol (Paris) 157:633-637

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743-753

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, et al. (2002) A human genome diversity cell line panel. Science 296:261-262

Dagvadorj A, Petersen RB, Lee HS, Cervenakova L, Shatunov A, Budka H, Brown P, Gambetti P, Goldfarb LG (2002) Spontaneous mutations in the prion protein gene causing transmissible spongiform encephalopathy. Ann Neurol 52:355-359

Goldfarb LG, Brown P, McCombie WR, Goldgaber D, Swergold GD, Wills PR, Cervenakova L, Baron H, Gibbs CJ, Jr., Gajdusek DC (1991) Transmissible familial Creutzfeldt-Jakob disease associated with five, seven, and eight extra octapeptide coding repeats in the PRNP gene. Proc Natl Acad Sci U S A 88:10926-10930

Goldman JS, Miller BL, Safar J, de Tourreil S, Martindale JL, Prusiner SB, Geschwind MD (2004) When sporadic disease is not sporadic: the potential for genetic etiology. Arch Neurol 61:213-216

Krawczak M, Ball EV, Cooper DN (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am J Hum Genet 63:474-488

Laplanche JL, Hachimi KH, Durieux I, Thuillet P, Defebvre L, Delasnerie-Laupretre N, Peoc'h K, Foncin JF, Destee A (1999) Prominent psychiatric features and early onset in an inherited prion disease with a new insertional mutation in the prion protein gene. Brain 122 (Pt 12):2375-2386

Lee IY, Westaway D, Smit AF, Wang K, Seto J, Chen L, Acharya C, Ankener M, Baskin D, Cooper C, Yao H, Prusiner SB, Hood LE (1998) Complete genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res 8:1022-1037

Mahal SP, Asante EA, Antoniou M, Collinge J (2001) Isolation and functional characterisation of the promoter region of the human prion protein gene. Gene 268:105-114

McCormack JE, Baybutt HN, Everington D, Will RG, Ironside JW, Manson JC (2002) PRNP contains both intronic and upstream regulatory regions that may influence susceptibility to Creutzfeldt-Jakob Disease. Gene 288:139-146

Mead S, Mahal SP, Beck J, Campbell T, Farrall M, Fisher E, Collinge J (2001) Sporadic--but not variant--Creutzfeldt-Jakob disease is associated with polymorphisms upstream of PRNP exon 1. Am J Hum Genet 69:1225-1235

Mead S, Stumpf MP, Whitfield J, Beck JA, Poulter M, Campbell T, Uphill JB, Goldstein D, Alpers M, Fisher EM, Collinge J (2003) Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. Science 300:640-643

Palmer MS, Dryden AJ, Hughes JT, Collinge J (1991) Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease. Nature 352:340-342

Rodriguez-Martínez A.B, Barreau C., Coupry I., Yagüe J., Sánchez-Valle R., Galdós-Alcelay L., Ibañez A., Digón A., Fernández-Manchola I., Goizet c., Castro A.,

Cuevas N., Alvarez-Alvarez M, de Pancorbo M.M., Arveiler B. and Zarranz J.J. (2005) Ancestral origins of the prion protein gene D178N mutation in the Basque Country. Human Genetics.

Sanchez-Valle R, Nos C, Yague J, Graus F, Dominguez A, Saiz A (2004) Clinical and genetic features of human prion diseases in Catalonia: 1993-2002. Eur J Neurol 11:649-655

Soldevila M, Calafell F, Andres AM, Yague J, Helgason A, Stefansson K, Bertranpetit J (2003) Prion susceptibility and protective alleles exhibit marked geographic differences. Hum Mutat 22:104-105

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162-1169

van Gool WA, Hensels GW, Hoogerwaard EM, Wiezer JH, Wesseling P, Bolhuis PA (1995) Hypokinesia and presenile dementia in a Dutch family with a novel insertion in the prion protein gene. Brain 118 (Pt 6):1565-1571

Windl O, Dempster M, Estibeiro JP, Lathe R, de Silva R, Esmonde T, Will R, Springbett A, Campbell TA, Sidle KC, Palmer MS, Collinge J (1996) Genetic basis of Creutzfeldt-Jakob disease in the United Kingdom: a systematic analysis of predisposing mutations and allelic variation in the PRNP gene. Hum Genet 98:259-264

Zarranz JJ, Digon A, Atares B, Arteagoitia JM, Carrera N, Fernandez-Manchola I, Fernandez-Martinez M, Fernandez-Maiztegui C, Forcadas I, Galdos L, Ibanez A, Lezcano E, Marti-Masso JF, Mendibe MM, Urtasun M, Uterga JM, Saracibar N, Velasco F, Gonzalez de Galdeano L (2005) Familial prion diseases in the Basque Country (Spain). Neuroepidemiology 24:103-109

RESULTS

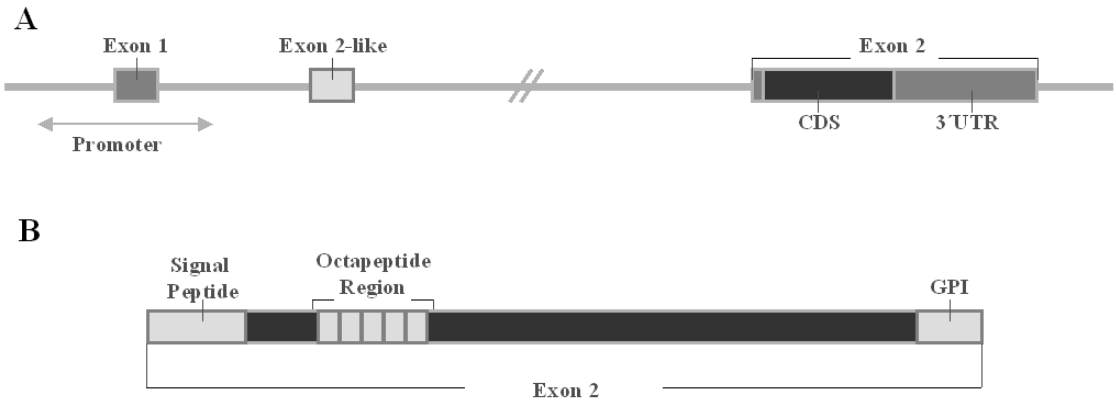
Table 1. Classification of patients with TSE		
FAMILIAL		
	200	10
	178	18
	OTHER	1
Total familial		**29**
SPORADIC		
	DEFINITE	32
	PROBABLE	19
	POSSIBLE	38
	OTHER	1
Total sporadic		**90**
TOTAL		**119**

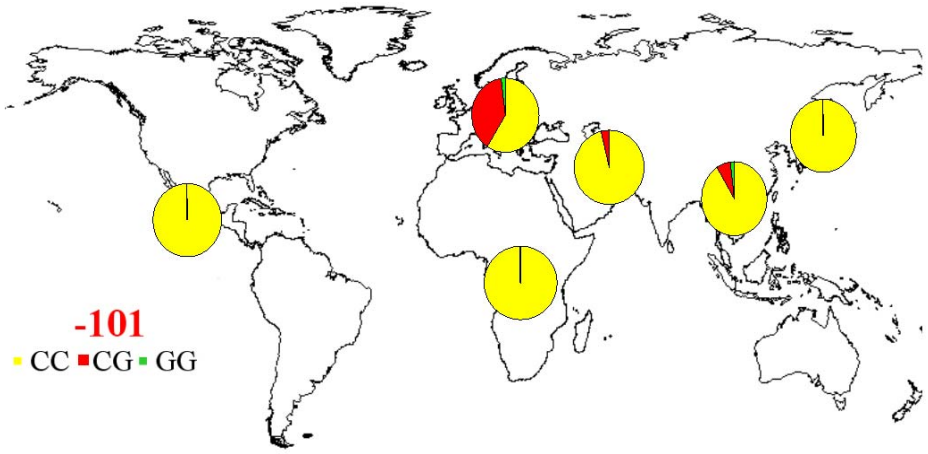
Table 2. Polymorphism in codon 129					
	N	**%**		**2N**	**%**
CJDs-D					
Met/Met	17	53.1	**Met**	42	65.6
Met/Val	8	25.0	**Val**	22	34.4
Val/Val	7	21.9			
Total	**32**			**64**	
CJDs-P					
Met/Met	10	52.6	**Met**	25	65.8
Met/Val	5	26.3	**Val**	13	34.2
Val/Val	4	21.1			
Total	**19**			**38**	
CJDs-PO					
Met/Met	19	50.0	**Met**	50	65.8
Met/Val	12	31.6	**Val**	26	34.2
Val/Val	7	18.4			
Total	**38**			**76**	
CJDs-D & P& OTHER					
Met/Met	28	53.8	**Met**	69	66.3
Met/Val	13	25.0	**Val**	35	33.7
Val/Val	11	21.2			
Total	**52**			**104**	
CJDf -200					
Met/Met	8	80.0	**Met**	18	90.0
Met/Val	2	20.0	**Val**	2	10.0
Val/Val	0	0.0			
Total	**10**			**20**	
FFI- 178					
Met/Met	12	66.7	**Met**	30	83.3
Met/Val	6	33.3	**Val**	6	16.7
Val/Val	0	0.0			
Total	**18**			**36**	
CJDf –200 & FFI 178 & OTHER					
Met/Met	20	69.0	**Met**	49	84.5
Met/Val	9	31.0	**Val**	9	15.5
Val/Val	0	0.0			
Total	**29**			**58**	

Table 3. Haplotypes found in the promoter region of *PRNP* in 69 patients

HAPL NAME	-563 G/A	-101 C/G	+242 G/T	+257 G/A	+310 G/C	SPORADIC D	SPORADIC P	SPORADIC PO	FAMILIAL 178	FAMILIAL 200	TOTAL CHR
P1	G	C	G	G	G	32	22	17	26	15	**112**
P2	G	C	G	G	C	2	2	2	2	1	**9**
P3	G	G	G	G	G	3	2	2	0	2	**9**
P4	A	C	G	G	G	1	1	3	0	0	**5**
P5	G	C	T	G	C	0	0	0	2	0	**2**
P6	G	C	G	A	G	0	1	0	0	0	**1**
						38	**28**	**24**	**30**	**18**	**138**

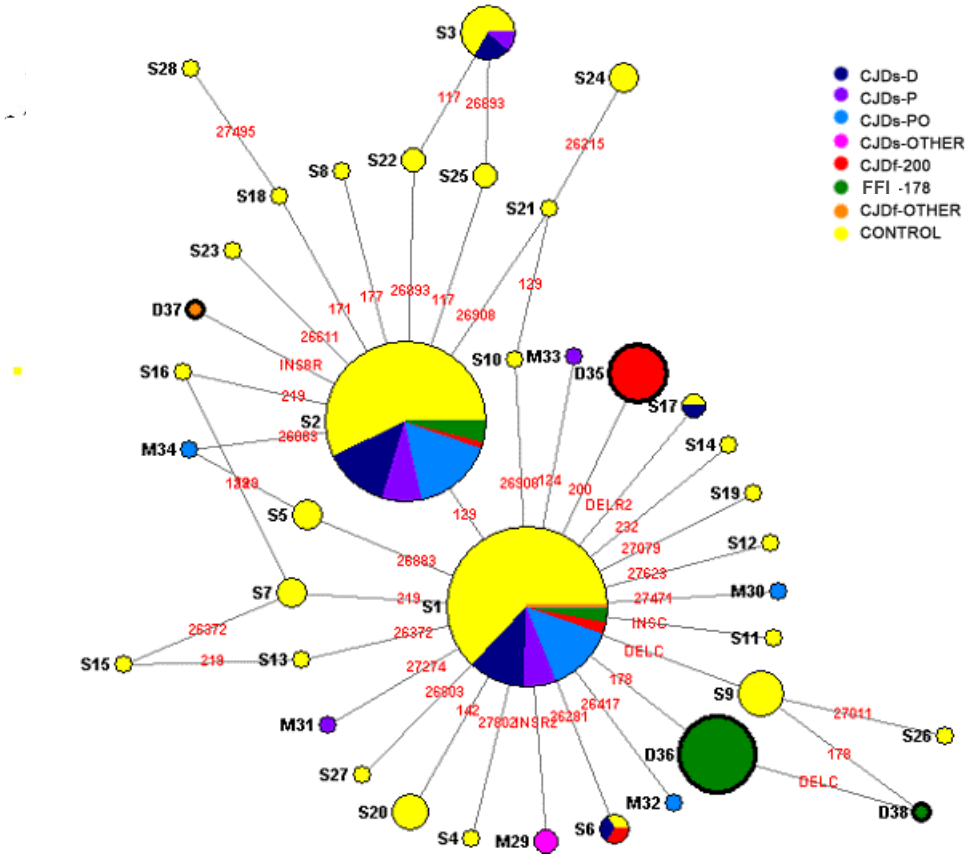
Table 4. Polymorphisms in the regulatory region.			
-563/Met 129	**CJDs D & P**	**CJDf**	**EUROPE**
G	53(98.1)	40 (100)	38 (95)
A	1 (1.9)	0 (0)	2 (5)
Fisher exact test	*Column 1 compared with column 2**: P = 0.575***		
	Column 1 compared with column 3: P= 0.388		
	Column 2 compared with column 3: P= 0.247		
Non Carriers -563	26 (96.3)	20 (100)	18 (90)
Carriers -563	1 (3.7)	0 (0)	2 (10)
Fisher exact test	*Column 1 compared with column 2: P= 0.575*		
	Column 1 compared with column 3: P= 0.387		
	Column 2 compared with column 3: P= 0.243		
-101/Met 129	**CJDs D & P**	**CJDf**	**EUROPE**
C	48 (88.9)	38 (95)	27 (67.5)
G	6 (11.1)	2 (5)	13 (32.5)
Fisher exact test	*Column 1 compared with column 2: P= 0.254*		
	Column 1 compared with column 3: P= 0.011		
	Column 2 compared with column 3: P< 0.002		
Non Carriers -101	21 (77.8)	18 (90)	8 (40)
Carriers -101	6 (22.2)	2 (10)	12 (60)
Fisher exact test	*Column 1 compared with column 2: P= 0.242*		
	Column 1 compared with column 3: P<0.010		
	*Column 2 compared with column 3: P**< 0.001***		
All Non Carriers	**20 (74.1)**	**18 (90)**	**7 (35)**
All Carriers	**7 (25.9)**	**2 (10)**	**13 (65)**
Fisher exact test	*Column 1 compared with column 2: P =0.159*		
	Column 1 compared with column 3: P= 0.008		
	Column 2 compared with column 3: P< 0.001		


A
Exon 1
Exon 2-like
Exon 2
CDS
3´UTR
Promoter
B
Signal Peptide
Octapeptide Region
GPI
Exon 2


-101
CC CG GG

RESULTS

	DIDII1111111122222222222222222222 ENENN12247770136666666667777777 LSLSS74921780922234688890024468 RRCC8 1871108901777920 22 R 5127133381941532	CONTROL WORLD	SPORADIC D	 P	 PO	 OTH	FAMILIAL 200	 178	 OTH	TOTAL
S1	A-A--ACAGACGGGTAGAATTCAACTGGTTG	220	40	23	48		7	10	1	**349**
S2	A-A--ACGGACGGGTAGAATTCAACTGGTTG	87	20	12	25		2	6		**152**
S3	A-A--GCGGACGGGTAGAATTCGACTGGTTG	6	2	1						**9**
S4	A-A--ACAGACGGGTAGAATTCAACTGGTTC	1								**1**
S5	A-A--ACAGACGGGTAGAATTGAACTGGTTG	2						1		**3**
S6	A-A--ACAGACGGGTAAAATTCAACTGGTTG	1	1				1			**31**
S7	A-A--ACAGACGGATAGAATTCAACTGGTTG	3								**3**
S8	A-A--ACGGATGGGTAGAATTCAACTGGTTG	1								**1**
S9	A----ACAGACGGGTAGAATTCAACTGGTTG	5						1		**6**
S10	A-A--ACAGACGGGTAGAATTCAGCTGGTTG	1								**1**
S11	A-AA-ACAGACGGGTAGAATTCAACTGGTTG	1								**1**
S12	A-A--ACAGACGGGTAGAATTCAACTGGTAG	1								**1**
S13	A-A--ACAGACGGGTAG-ATTCAACTGGTTG	1								**1**
S14	A-A--ACAGACGGGGAGAATTCAACTGGTTG	1								**1**
S15	A-A--ACAGACGGATAG-ATTCAACTGGTTG	1								**1**
S16	A-A--ACGGACGGATAGAATTCAACTGGTTG	1								**1**
S17	--A--ACAGACGGGTAGAATTCAACTGGTTG	1	1							**2**
S18	A-A--ACGGGCGGGTAGAATTCAACTGGTTG	1								**1**
S19	A-A--ACAGACGGGTAGAATTCAACGGGTTG	1								**1**
S20	A-A--ACAAACGGGTAGAATTCAACTGGTTG	4								**4**
S21	A-A--ACGGACGGGTAGAATTCAGCTGGTTG	1								**1**
S22	A-A--ACGGACGGGTAGAATTCGACTGGTTG	2								**2**
S23	A-A--ACGGACGGGTAGAACTCAACTGGTTG	1								**1**
S24	A-A--ACGGACGGGTTGAATTCAGCTGGTTG	3								**3**
S25	A-A--GCGGACGGGTAGAATTCAACTGGTTG	2								**2**
S26	A----ACAGACGGGTAGAATTCAATTGGTTG	1								**1**
S27	A-A--ACAGACGGGTAGAATCCAACTGGTTG	1								**1**
S28	A-A--ACGGGCGGGTAGAATTCAACTGG-TG	1								**1**
M29	AAA--ACAGACGGGTAGAATTCAACTGGTTG					2				**2**
M30	A-A--ACAGACGGGTAGAATTCAACTGATTG				1					**1**
M31	A-A--ACAGACGGGTAGAATTCAACTAGTTG			1						**1**
M32	A-A--ACAGACGGGTAGAGTTCAACTGGTTG				1					**1**
M33	A-A--AGAGACGGGTAGAATTCAACTGGTTG			1						**1**
M34	A-A--ACGGACGGGTAGAATTGAACTGGTTG				1					**1**
D35	A-A--ACAGACGAGTAGAATTCAACTGGTTG						10			**10**
D36	A-A--ACAGACAGGTAGAATTCAACTGGTTG							17		**17**
D37	A-A-AACGGACGGGTAGAATTCAACTGGTTG								1	**1**
D38	A----ACAGACAGGTAGAATTCAACTGGTTG							1		**1**
TOTAL		**348**	**64**	**38**	**76**	**2**	**20**	**36**	**2**	**586**


CJDs-D
CJDs-P
CJDs-PO
CJDs-OTHER
CJDf-200
FFI -178
CJDf-OTHER
CONTROL
S3
S28
S24
S22
S8
S25
S18
S21
S23
D37
S16
S10
M33
D35
S17
S2
M34
S14
S19
S5
S12
S7
M30
S1
S11
S15
S13
S9
M31
S26
D36
S27
M32
S20
D38
S4
M29
S6
27495
26215
26908
INS8R
249
26883
DELR2
232
27079
27623
27471
INSC
DELC
178
27011
26372
219
27274
26803
142
27802NSR2
26281
26417

RESULTS

Figure 1. A).Human *PRNP* gene. B) Exon 2 of the PRNP gene. *CDS, UTR* and *GPI* stand for coding sequence, untranslated region and glycosylphophatidylinositol anchor, respectively

Figure 2. Genotype frequencies for –101 polymorphism.

Figure 3. Haplotypes found in the exon 2 region of *PRNP* gene. Names of each haplotype are shown on the left. Polymorphic positions found in this region are shown in the top. Control world is referred to the HGDP-CEPH samples used in Soldevila et al. (unpublished results). Absolute frequencies are shown for each group.

Figure 4. Median Joining network for the haplotypes found in exon 2. In yellow, termed `controls´, haplotypes found in samples of the HGPD-CEPH panel used in Soldevila et al. (unpublished results). In bold circles, chromosomes that carry a familial mutation, 178 in green, or 200 in red. The other green and red circles belong to chromosomes that carry patients suffering familial prion diseases, but don´t have any causative

6. DISCUSSION

On the polymorphism on codon 129 and 219 from the prion protein gene in humans

Prion protein gene or PRNP has been analysed in different human populations in order to elucidate the extent and pattern of variation in this gene. The variation at the DNA level is useful from both an evolutionary and a medical point of view.

The prion scientific community has focussed their attention on a specific position within the PRNP gene. A polymorphism in codon 129 has been related to an increased susceptibility for developing prion diseases. Prion diseases are classified as sporadic, familial or acquired. All types of prion diseases are influenced by this polymorphism. Homozygotes, significantly for Methionine allele in codon 129, have been found in a higher ratio than heterozygotes in patients suffering acquired or sporadic prion diseases. Variant of Creutzfeldt-Jakob disease (vCJD) is classified as acquired, and the ratio of Methionine or Valine allele is striking: all the cases are homozygotes for Methionine allele in codon 129. Familial cases are influenced by this polymorphism, because an association between this codon 129 and the polymorphism in codon 178 could give different pathologies, either FFI or CJD depending on the combination of these variants. Thus, this polymorphism is essential to understand this pathology.

In **Chapter II**, we have seen that this common PRNP gene polymorphism is critical in the susceptibility to all types of prion diseases and what we have seen

is that there are big differences in the frequency of both alleles, Methionine and Valine, in different populations. Especially, frequencies in Native Americans and in Asiatic populations are very different and thus, further epidemiological studies need to be done to clarify if the prevalence of these prion diseases are the same in these two regions, as it is currently held, however little epidemiological studies have been done in these regions. Another possibility is that another polymorphism such as the one in codon 219 could somehow *compensate* this susceptibility. As we have observed, there is a protective allele in this position (Lysine) that is mainly present in Asiatic populations, and thus, maybe it is *compensating* –totally or partially-, the effect of codon 129.

At the same time, an efficient new approach was proposed for the simultaneous detection of either four or five sequence variants, including codon 129 and codon 219. The other variants included in the test were related to familial prion diseases, such as Creutzfeldt-Jakob, Fatal Familial Insomnia and Gerstmann-Sträussler-Scheinker.

On the polymorphism on codon 129 and 219 from the prion protein gene in chimpanzees

As we have seen previously, polymorphisms in codon 129 and 219 in humans are key factors in the development of prion diseases. For this reason we decided to study the variation in PRNP gene in our closest species, chimpanzees. The analysis of the differences between humans and

chimpanzees at the DNA level, which is called comparative genetics, help us to understand the species-specific characteristics. Just a limited number of chimpanzee intraespecific studies have been done because of the problems to obtain a big and representative sample of *Pan troglodytes*. The objective of this study (**Chapter I**) was to analyse the coding region in order to see if these two important polymorphisms in humans were also polymorphic in this species.

We sequence part of the coding region of the PRNP gene in sixty-five chimpanzee samples and our results showed that these polymorphic positions in humans are fixed in chimpanzees. That means that *a priori*, no susceptibility differences exist among chimpanzees, as all of them have the same allele and so, this distinguishes chimpanzees and humans in this case, because humans have more than one allele. Moreover, the allele fixed in position 129 and 219 are respectively, Methionine and Glycine, where Methionine is related with an increased susceptibility to the development of prion diseases, and Glycine in position 219 is associated with a protective effect. This means that *a priori*, chimpanzees are more susceptible to develop prion diseases than humans from a genetic point of view.

Different studies demonstrate that this close species can develop prion diseases. This was studied *in vitro* and *in vivo* (Gajdusek and Gibbs 1975; Cervenakova et al. 1994; Bons et al. 1999), and it was demonstrated that primates can develop prion protein diseases, and that also, they were infected also by mad cow disease epidemia (captive primates in zoological parks).

However, it is important to notice that even though they can develop prion diseases, no sporadic or familial case have been detected in chimpanzees. This

may be for many different reasons, but one possible explanation, could be that the life span of chimpanzees is shorter than that of humans. The age at onset of prion diseases is usually in the 45-75 year age group. This can explain why human dementias are not detected in chimpanzees, but also, and in the same direction, there is a lack of studies about chimpanzees' diseases, that could be undetecting these pathologies.

In the same study (**Chapter I**), we describe a novel non-synonymous polymorphism that is exclusive to chimpanzees, that is located in codon 148, a conserved region in the gene. This variation could play a role in prion structure and for this reason we decided to analyse the effect of this change. The functional impact on the protein function of the Arginine to Histidine change in position 148 would be limited, according to the accepted structural model. Moreover, no symptomatology was observed in the female chimpanzee that had the mutation, so we proposed that this variant is non-pathogenic. However, other factors (differences in life span, species characteristics and other susceptibility or protective alleles) could influence the outcome of this polymorphism.

Another interesting point was to check if mutations in the prion protein gene could influence the results in the widely used PRIONICS test. This test is in use in most of the countries to detect bovine spongiform encephalopathy (BSE), commonly referred to as "mad cow disease", and it is based on a simple western-blot procedure. The non-synonymous change found in chimpanzees was located in a conservative region where most of antibodies have been designed to interact with. Antibodies used in PRIONICS test recognize this

region. We used synthetic peptides (as only DNA from this sample was available) to check if there was a failure in the detection of the abnormal peptide, thus generating negative results if the variant was disease causative. This highlighted the importance in knowing all the variants in the species that suffer prion diseases, as detection tests, such as PRIONICS, could fail in some cases due to genetic variants.

On the methods to detect selection

Different methods can be used to detect selection based on sequencing data, as we have seen in the introduction. Among them one of the most widely used test is Tajima´s D, together with Fu and Li´s tests. Mead et al. (Mead et al. 2003) analysed the prion protein gene in a sample representing the variation in human species and found that balancing selection was driving the evolution of this gene. However some problems with the methodology used were underlined in another study by (Kreitman and Di Rienzo 2004). They suggested that the strategy followed could have biased the results by selectively excluding low-frequency variants and thus affecting the main conclusions. In **Chapter III** we present interesting experimental data that confirm their suggestions, consisting of the results of a full resequencing of 2378 bp from the human *PRNP* gene, corresponding to the entire exon 2, in 174 humans. Our main conclusions are that the method used can introduce important biases in the analysis, and that tests like Tajima´s D are not suitable for SNP data. Moreover, our data present

overwhelming evidence against the hypothesis of repeated balancing selection events in human history affecting the PRNP gene, as was previously suggested. Thus, in short, we dismissed the claims that cannibalism had had a main role in human history, as the proposed evidence cannot be maintained.

The topic of ascertainment bias had been previously brought up and investigated in a number of other papers (Nielsen 2000; Wakeley et al. 2001; Clark et al. 2003; Nielsen and Signorovitch 2003; Nielsen et al. 2004), and they show that the design of the study and the way the data is analysed are very important in the avoidance of incorrect conclusions.

Sequencing studies, although time consuming and very expensive, are necessary to allow us to apply the classical neutrality tests and detect the selective forces that are shaping the pattern of variation in our genes.

On the evolution gene

In **Chapter IV** we studied the patterns of nucleotide variation associated with PRNP exon 2. As prion protein has some special characteristics that make it unique, it was interesting to investigate the evolutionary history of this gene. A previous study has claimed that balancing selection is acting on this gene, but it has been controversial because of the bias in the method used to obtain the polymorphisms (Mead et al. 2003) as we have seen in **Chapter III.** In our study we found no evidence of ancient and recurrent balancing selection. However, when neutrality tests were applied to our data, positive values of

Tajima´s D were obtained, indicating either positive or purifying selection. Similar conclusions were found when other tests and other analyses were applied. Whereas the low number of non-synonymous variants supports the action of purifying selection, one study by (Gordo et al. 2002), indicates that extremely negative values of Tajima's D can only be obtained with positive selection. In the light of the results obtained, positive selection appears to be the general force operating on this PRNP gene and in any case recurrent balancing selection cannot be accepted.

Some interesting features can be seen when analysing different continental regions. Tajima´s D values were negative (but non-significant) in all the regions except for Americas, where due to the short variation found in this region the test could not be applied. This limited genetic variation (only the 129 polymorphism was found) could be explained by selection forces acting on this PRNP gene, such as selective sweeps favouring the increase of frequency of Methionine or specially Valine in the polymorphic 129 position. Further studies should be done in Americas to confirm this hypothesis.

In prion protein gene, the history of the evolution of this gene is not simple, but positive selection seems to be the main force acting on it. However, in some local regions, like in the Fore or in the Americas region, some particular pressures may be acting, thus, explaining some of the specific features found. This is especially interesting in the light of the fact that cannibalism has been shown to be operating in these two regions. Kuru (a prion disease, see introduction) was finally eradicated when ritual cannibalistic practices were

banned. In the Americas, ancient cannibalism practices have been better documented than elsewhere (confirmed in Anasazi (Marlar et al. 2000) and suggested in other groups (Gibbons 1997) (Defleur et al. 1999), this may have driven the Val allele frequencies upwards, as has been suggested for the similarly Val-rich Fore population (Mead et al. 2003).

A complex story, including selective forces acting in different regions, and at specific times in human history, seem to be a plausible scenario for the human *PRNP* gene.

On the patients suffering sporadic prion diseases

As we have seen in the introduction, human prion diseases are classified as familial, acquired or sporadic. Even if sporadic cases represent about 80-85% of the total cases, little is known about the reasons that cause a person to go on to develop a prion disease without any associated mutation in the PRNP gene.
We suggest different reasons to explain the high ratio of sporadic cases:

1. *Some cases classified as sporadic are not related with prion diseases*: Some other dementias, such as Alzheimer, can sometimes give similar phenotypes. For example this has been suggested in the case of Huntington disease. Although this may be the case in possible and probable cases of sporadic prion disease, it does not account for any of the definite sporadic cases.

2. *Some cases classified as sporadic may be acquired*: It has been seen that whereas it was thought that BSE only caused a new form of the disease called variant CJD (vCJD), a study in mice from a team led by John Collinge suggests that it may also cause a disease indistinguishable from the commonest form of classical, or 'sporadic', CJD (Asante et al. 2002). If the group's mouse model is relevant to the human disease, the results also suggest that the true extent of infection may be difficult to assess because of the large number of asymptomatic carriers. In their experiments, they have shown that the pattern in the western blot of some BSE infected mouse was consistent with sporadic CJD. The number of cases of sporadic CJD has been rising in Britain since the 1970s, and this had been attributed to better monitoring for the condition. After these studies, researchers led by Adriano Aguzzi reported a sudden increase in sporadic CJD figures in Switzerland in 2001, and suggested that infection with BSE might be to blame (Glatzel et al. 2002). Collinge's data provided worrying molecular evidence that BSE might be to blame for the rise in sporadic CJD. However, if it is true, it is important to point out that only a low ratio of sporadic cases could be acquired, but still, the main question of why a person develops a sporadic prion disease is not well understood.
3. *Some cases classified as sporadic may be genetic:* This may be due to a number of reasons:
    a. *Lack of genetic analysis:* It has been estimated that 10% of apparently sporadic TSE cases show the presence of a *PRNP*

mutation after a detailed resequencing study of the coding part (cds) of the cases (The EUROCJD Group. (2001). (See for example (Goldman et al. 2004)

b. *De novo* mutations: explain the lack of family history in some cases that were classified as sporadic and it can be explained by the presence of some CpG dinucleotides, such as mutation at codon 178 (causing TSE) (Windl et al. 1996; Dagvadorj et al. 2002), However this represents a low frequency of cases. This could be solved by sequencing the cds.

c. *Lack of family history may classify the case as sporadic*: This can be resolved by sequencing the cds. Recent data reports that a large percentage of genetic case do not have a positive family history.

d. *Regions other than coding part can be important*: Regions such as the promoter or the rest of exon 2 (5´UTR) may have an affect on the susceptibility to human prion diseases. These regions are not studied in the normal screening for mutations, and thus, the effect of them in the misclassification of sporadic and genetic cases could be very important.

In **Chapter V** we tried elucidate if hypothesis d is plausible by analysing the promoter and the complete exon 2 in some patients suffering prion diseases. We knew in advance that in the promoter region some polymorphisms

(McCormack et al. 2002) were related with the increased susceptibility to develop sporadic prion disease, together with the polymorphism in codon 129.

In our project we analyse a total of 119 patients with prion diseases, including familial and sporadic cases.

No mutation in exon 2 was major associated with sporadic CJD, and thus, sporadic cases have no particular polymorphism in the 5´UTR of the PRNP gene that requires testing. However, some low frequency mutations appear only in some sporadic cases. This can be interpreted in two ways: first, that they are neutral polymorphisms with a low frequency and they could also be found in controls with a bigger sample, or second, that some low frequency variants can explain a low proportion of sporadic cases, and so, these cases could become genetic or familial. We can't discard the possibility that some polymorphisms in the 5´UTR region can alter the stability of the mRNA, and thus having an affect on the protein folding, but even if this could indeed be true, most of the sporadic cases cannot be explained by this hypothesis.

In the same study we decided to focus on the promoter of the prion gene. It is very well-known that many polymorphisms in the promoters of different genes have an extraordinary effect in the expression of their proteins, and so, many of them have been described as being causative for disease. In the PRNP gene, what we have seen is that being a carrier of more than one rare allele in the promoter increases the susceptibility to develop sporadic CJD when it is coupled with Methionine allele in codon 129. The increased risk associated with carrying

a rare allele may be due to a more complex model where two or more alleles independently increase susceptibility to sporadic CJD (McCormack et al. 2002). But it is very difficult to prove that such complex combinations are indeed causal factors.

It has been well established that polymorphisms in the coding region of PRNP are associated with human TSEs but the polymorphisms described here are the first loci outside the prion coding region that may influence any form of CJD. Alterations in susceptibility to TSE disease are currently hypothesized to be linked to structural differences in the PrP protein structure, which alter the efficiency of replication of the infectious agent. However, alterations in TSE susceptibility could also be modulated by the expression levels of PrP.

Our preliminary data indicates that in the promoter of PRNP gene interesting polymorphisms exist that should be taken into consideration in further analysis of patients with sporadic CJD.

General

In conclusion we can say that population genetics can help us to understand the pattern of genetic variation and the forces acting on an specific gene, such as PRNP, but may also aid in the comprehension of specific features and give

reasons why differences between humans and chimpanzees exist and also, if there are differences among humans.

In the prion gene, we have observed that the particular pattern of variation that we have found in this gene based on sequencing data of samples representing the global human variation is due to positive selection, and that the method and the approach used to detect this selection critical. Ascertainment bias can be introduced by using SNP data and applying neutrality tests based on sequence diversity, therefore leading to anomalous conclusions being drawn.

Moreover, we have seen that there are specific positions in the gene, such as a polymorphism in codon 129 and 219 that can explain where the selection is specifically acting. These polymorphisms are important from a medical point of view, as depending on which of the two alleles are present, there is more or less susceptibility to or even protection against prion diseases. This means that among humans there are individuals more susceptible to develop prion diseases than others. We have seen that these positions are fixed in chimpanzees, so in these specific variants chimpanzees have no differences, and moreover, the allele fixed is the one that gives an increase susceptibility to these diseases in humans. But polymorphisms that are present in humans are not normally shared with chimpanzees, and in this sense, we found other polymorphisms in this close species that were not found in our species. The functional impact of these positions should be checked in more detail, but our studies indicate that the aminoacid change that we found in position 148 has a low impact on the structure, and so this can represent a neutral polymorphism.

The normal variation that we found in controls can be then compared with patients for the same region and so attempting to distinguish some relevant positions that can differentiate both groups. Although no specific difference was found to be generally present in patients for the 3´UTR region of the gene, some polymorphisms in the promoter indicate that there could be interesting positions that will need to be taken into consideration for further genetic studies of patients suffering sporadic CJD.

The resequencing of PRNP in a very large sample of humans and chimpanzees has provided a great deal of information on this gene. The amount of data generated is striking, and very promising. Our work has therefore been essential to clarify most of the evolutionary aspects of the PRNP gene. Moreover, this data has provided insights into a better understanding of prion diseases from a genetic point of view which could lead to new possible ways of treating these fatal disorders in the future.

7. GLOSSARY

AMYLOID: A generic term that describes fibrillar aggregates organized in a β-pleated-sheet conformation. These aggregates exhibit specific tinctorial properties, including the ability to emit a green birefringent glow after staining with the dye Congo red, and the capacity to bind the fluorochrome thioflavin S. There are >12 human diseases of different etiology characterized by the extracellular deposition of amyloid. The amyloid fibrils are usually composed of proteolytic fragments of normal or mutant gene products. So far, there are over 16 different proteins involved in amyloid deposition in distinct tissues.

CONFORMATIONAL TRANSMISSION: The process by which the three-dimensional structure of one protein molecule is transferred to another molecule that acquires the structural and functional properties of the former molecule.

CHAPERONE PROTEIN: A group of molecules that assist diverse proteins to acquire their native folding. Synthetic mini-chaperone peptides refer to the concept of short peptides engineered to interact with and alter the folding of specific proteins.

PRION: A causative agent of transmissible spongiform encephalopathy (TSE), with unconventional properties. The term does not have structural implications other than that a protein is an essential component.

PROTEIN-ONLY HYPOTHESIS: Maintains that the prion is devoid of informational nucleic acid, and that the essential pathogenic component is protein (or glycoprotein). Genetic evidence indicates that the protein is an abnormal form of PrP (perhaps identical with PrPSc).Association with other

'non-informational' molecules (such as lipids or glycosamino glycans) is not excluded.

PRPC: PrPC is the naturally occurring form of the mature *Prnp* gene product. Its presence in a given cell type is necessary, but not sufficient, for replication of the prion.

PRPSC: PrPSc, found in TSE sufferers, is an 'abnormal' form of the mature PrPC protein, defined as being partly resistant to digestion by proteinase K under standardized conditions. It is believed to differ from PrPC only (or mainly) conformationally, and is often considered to be the transmissible agent or prion.

PROTEIN MISFOLDING: Alteration in the secondary and tertiary structure of a protein, resulting in a protein with altered conformation.

SPONGIFORM DEGENERATION: The most typical pathological feature of the brain of individuals affected by TSEs, consisting of the vacuolation of brain tissue, which results in holes in the brain.

B**-SHEET:** β-sheets are one of the prevalent, repetitive secondary structural motifs in folded proteins. β-sheets are formed from alternating peptide pleated strands linked by hydrogen bonding between the NH and CO groups of the peptidic bond. In β-sheets, the β-strands can come from a different region of the same protein or from a different molecule. Therefore, formation of β-sheets can be stabilized by protein oligomerization or aggregation.

8. REFERENCES

(1998) Human transmissible spongiform encephalopathies. Wkly Epidemiol Rec 73:361-365

(2001) Genetic epidemiology of Creutzfeldt-Jakob disease in Europe. Rev Neurol (Paris) 157:633-637

Aguzzi A, Brandner S (1999) The genetics of prions--a contradiction in terms? Lancet 354 Suppl 1:SI22-25

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2:e286

Aldhous P (1990) BSE: spongiform encephalopathy found in cat. Nature 345:194

Alper T, Cramp WA, Haig DA, Clarke MC (1967) Does the agent of scrapie replicate without nucleic acid? Nature 214:764-766

Alper T, Haig DA, Clarke MC (1966) The exceptionally small size of the scrapie agent. Biochem Biophys Res Commun 22:278-284

Alperovitch A, Zerr I, Pocchiari M, Mitrova E, de Pedro Cuesta J, Hegyi I, Collins S, Kretzschmar H, van Duijn C, Will RG (1999) Codon 129 prion protein genotype and sporadic Creutzfeldt-Jakob disease. Lancet 353:1673-1674

Alpers M, Rail L (1971) Kuru and Creutzfeldt-Jakob disease: clinical and aetiological aspects. Proc Aust Assoc Neurol 8:7-15

Andres AM, Soldevila M, Saitou N, Volpini V, Calafell F, Bertranpetit J (2003) Understanding the dynamics of Spinocerebellar Ataxia 8 (SCA8) locus through a comparative genetic approach in humans and apes. Neurosci Lett 336:143-146

Andrews NJ, Farrington CP, Ward HJ, Cousens SN, Smith PG, Molesworth AM, Knight RS, Ironside JW, Will RG (2003) Deaths from variant Creutzfeldt-Jakob disease in the UK. Lancet 361:751-752

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223-230

Arnason U, Gullberg A, Burguete AS, Janke A (2000) Molecular estimates of primate divergences and new hypotheses for primate dispersal and the origin of modern humans. Hereditas 133:217-228

Asante EA, Linehan JM, Desbruslais M, Joiner S, Gowland I, Wood AL, Welch J, Hill AF, Lloyd SE, Wadsworth JD, Collinge J (2002) BSE prions propagate as either variant CJD-like or sporadic CJD-like prion strains in transgenic mice expressing human prion protein. Embo J 21:6358-6366

Aucouturier P, Kascsak RJ, Frangione B, Wisniewski T (1999) Biochemical and conformational variability of human prion strains in sporadic Creutzfeldt-Jakob disease. Neurosci Lett 274:33-36

Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99-111

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743-753

Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL (1997) An apportionment of human DNA diversity. Proc Natl Acad Sci U S A 94:4516-4519

Bertoni JM, Label LS, Sackelleres JC, Hicks SP (1983) Supranuclear gaze palsy in familial Creutzfeldt-Jakob disease. Arch Neurol 40:618-622

Bessen RA, Marsh RF (1994) Distinct PrP properties suggest the molecular basis of strain variation in transmissible mink encephalopathy. J Virol 68:7859-7868

Bolton DC, McKinley MP, Prusiner SB (1982) Identification of a protein that purifies with the scrapie prion. Science 218:1309-1311

Bons N, Mestre-Frances N, Belli P, Cathala F, Gajdusek DC, Brown P (1999) Natural and experimental oral infection of nonhuman primates by bovine spongiform encephalopathy agents. Proc Natl Acad Sci U S A 96:4046-4051

Brown P (1997) The risk of bovine spongiform encephalopathy ('mad cow disease') to human health. Jama 278:1008-1011

REFERENCES

Brown P, Bradley R (1998) 1755 and all that: a historical primer of transmissible spongiform encephalopathy. Bmj 317:1688-1692

Brown P, Cathala F, Castaigne P, Gajdusek DC (1986) Creutzfeldt-Jakob disease: clinical analysis of a consecutive series of 230 neuropathologically verified cases. Ann Neurol 20:597-602

Brown P, Cervenakova L, Goldfarb LG, McCombie WR, Rubenstein R, Will RG, Pocchiari M, Martinez-Lage JF, Scalici C, Masullo C, et al. (1994) Iatrogenic Creutzfeldt-Jakob disease: an example of the interplay between ancient genes and modern medicine. Neurology 44:291-293

Brown P, Galvez S, Goldfarb LG, Nieto A, Cartier L, Gibbs CJ, Jr., Gajdusek DC (1992) Familial Creutzfeldt-Jakob disease in Chile is associated with the codon 200 mutation of the PRNP amyloid precursor gene on chromosome 20. J Neurol Sci 112:65-67

Brown P, Goldfarb LG, Gajdusek DC (1991) The new biology of spongiform encephalopathy: infectious amyloidoses with a genetic twist. Lancet 337:1019-1022

Brown P, Preece M, Brandel JP, Sato T, McShane L, Zerr I, Fletcher A, Will RG, Pocchiari M, Cashman NR, d'Aignaux JH, Cervenakova L, Fradkin J, Schonberger LB, Collins SJ (2000) Iatrogenic Creutzfeldt-Jakob disease at the millennium. Neurology 55:1075-1081

Brown P, Rodgers-Johnson P, Cathala F, Gibbs CJ, Jr., Gajdusek DC (1984) Creutzfeldt-Jakob disease of long duration: clinicopathological characteristics, transmissibility, and differential diagnosis. Ann Neurol 16:295-304

Brown P, Will RG, Bradley R, Asher DM, Detwiler L (2001) Bovine spongiform encephalopathy and variant Creutzfeldt-Jakob disease: background, evolution, and current concerns. Emerg Infect Dis 7:6-16

Bruce ME, Will RG, Ironside JW, McConnell I, Drummond D, Suttie A, McCardle L, Chree A, Hope J, Birkett C, Cousens S, Fraser H, Bostock CJ (1997) Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent. Nature 389:498-501

Bueler H, Aguzzi A, Sailer A, Greiner RA, Autenried P, Aguet M, Weissmann C (1993) Mice devoid of PrP are resistant to scrapie. Cell 73:1339-1347

Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, Barlow C (2003) Elevated gene expression levels distinguish human from non-human primate brains. Proc Natl Acad Sci U S A 100:13030-13035

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, et al. (2002) A human genome diversity cell line panel. Science 296:261-262

REFERENCES

Carp RI, Kascsak RJ, Rubenstein R, Merz PA (1994) The puzzle of PrPSc and infectivity--do the pieces fit? Trends Neurosci 17:148-149

Carroll SB (2003) Genetics and the making of Homo sapiens. Nature 422:849-857

Caughey B, Raymond GJ, Bessen RA (1998) Strain-dependent differences in beta-sheet conformations of abnormal prion protein. J Biol Chem 273:32230-32235

Cervenakova L, Brown P, Goldfarb LG, Nagle J, Pettrone K, Rubenstein R, Dubnick M, Gibbs CJ, Jr., Gajdusek DC (1994) Infectious amyloid precursor gene sequences in primates used for experimental transmission of human spongiform encephalopathy. Proc Natl Acad Sci U S A 91:12159-12162

Cervenakova L, Goldfarb LG, Garruto R, Lee HS, Gajdusek DC, Brown P (1998) Phenotype-genotype studies in kuru: implications for new variant Creutzfeldt-Jakob disease. Proc Natl Acad Sci U S A 95:13239-13241

Chabry J, Caughey B, Chesebro B (1998) Specific inhibition of in vitro formation of protease-resistant prion protein by synthetic peptides. J Biol Chem 273:13203-13207

Chen FC, Vallender EJ, Wang H, Tzeng CS, Li WH (2001) Genomic divergence between human and chimpanzee estimated from large-scale alignments of genomic sequences. J Hered 92:481-489

Chesebro B (1998) BSE and prions: uncertainties about the agent. Science 279:42-43

Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, Wright KL, Muchmore EA, Nelson DL, Warren ST, Varki A (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. Proc Natl Acad Sci U S A 95:11751-11756

Clark AG, Nielsen R, Signorovitch J, Matise TC, Glanowski S, Heil J, Winn-Deen ES, Holden AL, Lai E (2003) Linkage disequilibrium and inference of ancestral recombination in 538 single-nucleotide polymorphism clusters across the human genome. Am J Hum Genet 73:285-300

Cohen FE, Prusiner SB (1998) Pathologic conformations of prion proteins. Annu Rev Biochem 67:793-819

Collee JG, Bradley R (1997a) BSE: a decade on--Part 2. Lancet 349:715-721

Collee JG, Bradley R (1997b) BSE: a decade on--Part I. Lancet 349:636-641

Collinge J, Sidle KC, Meads J, Ironside J, Hill AF (1996) Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. Nature 383:685-690

Cooper G, Rubinsztein DC, Amos W (1998) Ascertainment bias cannot entirely account for human microsatellites being longer than their chimpanzee homologues. Hum Mol Genet 7:1425-1429

Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, Stephens M (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat Genet 36:700-706

Crouau-Roy B, Bouzekri N, Carcassi C, Clayton J, Contu L, Cambon-Thomsen A (1996) Strong association between microsatellites and an HLA-B, DR haplotype (B18-DR3): implication for microsatellite evolution. Immunogenetics 43:255-260

Crow TJ (2002) Handedness, language lateralisation and anatomical asymmetry: relevance of protocadherin XY to hominid speciation and the aetiology of psychosis. Point of view. Br J Psychiatry 181:295-297

Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex. Am J Hum Genet 71:759-776

Dagvadorj A, Petersen RB, Lee HS, Cervenakova L, Shatunov A, Budka H, Brown P, Gambetti P, Goldfarb LG (2002) Spontaneous mutations in the prion protein gene causing transmissible spongiform encephalopathy. Ann Neurol 52:355-359

De Gioia L, Selvaggini C, Ghibaudi E, Diomede L, Bugiani O, Forloni G, Tagliavini F, Salmona M (1994) Conformational polymorphism of the amyloidogenic and neurotoxic peptide homologous to residues 106-126 of the prion protein. J Biol Chem 269:7859-7862

Defleur A, White T, Valensi P, Slimak L, Cregut-Bonnoure E (1999) Neanderthal cannibalism at Moula-Guercy, Ardeche, France. Science 286:128-131

Dormont D (1999) Agents that cause transmissible subacute spongiform encephalopathies. Biomed Pharmacother 53:3-8

Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, Lahn BT (2004) Accelerated evolution of nervous system genes in the origin of Homo sapiens. Cell 119:1027-1040

Eanes WF, Kirchner M, Yoon J (1993) Evidence for adaptive evolution of the G6pd gene in the Drosophila melanogaster and Drosophila simulans lineages. Proc Natl Acad Sci U S A 90:7475-7479

Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Paabo S (2002) Intra- and interspecific variation in primate gene expression patterns. Science 296:340-343

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3:87-112

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368-376

REFERENCES

Ferris SD, Wilson AC, Brown WM (1981) Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc Natl Acad Sci U S A 78:2432-2436

Fleetwood AJ, Furley CW (1990) Spongiform encephalopathy in an eland. Vet Rec 126:408-409

Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, Meltesen L, Brenton M, Hink R, Burgers S, Hernandez-Boussard T, Karimpour-Fard A, Glueck D, McGavran L, Berry R, Pollack J, Sikela JM (2004) Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol 2:E207

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693-709

Gabizon R, McKinley MP, Groth D, Prusiner SB (1988) Immunoaffinity purification and neutralization of scrapie prion infectivity. Proc Natl Acad Sci U S A 85:6617-6621

Gagneux P, Amess B, Diaz S, Moore S, Patel T, Dillmann W, Parekh R, Varki A (2001) Proteomic comparison of human and great ape blood plasma reveals conserved glycosylation and differences in thyroid hormone metabolism. Am J Phys Anthropol 115:99-109

Gagneux P, Wills C, Gerloff U, Tautz D, Morin PA, Boesch C, Fruth B, Hohmann G, Ryder OA, Woodruff DS (1999) Mitochondrial sequences show diverse

evolutionary histories of African hominoids. Proc Natl Acad Sci U S A 96:5077-5082

Gajdusek DC, Gibbs CJ, Jr. (1975) Familial and sporadic chronic neurological degenerative disorders transmitted from man to primates. Adv Neurol 10:291-317

Gajdusek DC, Zigas V (1957) Degenerative disease of the central nervous system in New Guinea; the endemic occurrence of kuru in the native population. N Engl J Med 257:974-978

Galtier N, Depaulis F, Barton NH (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics 155:981-987

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Mol Biol Evol 12:594-603

Gibbons A (1997) Archaeologists rediscover cannibals. Science 277:635-637

Gilad Y, Man O, Paabo S, Lancet D (2003) Human specific loss of olfactory receptor genes. Proc Natl Acad Sci U S A 100:3324-3327

Glatzel M, Rogivue C, Ghani A, Streffer JR, Amsler L, Aguzzi A (2002) Incidence of Creutzfeldt-Jakob disease in Switzerland. Lancet 360:139-141

Glazko GV, Nei M (2003) Estimation of divergence times for major lineages of primate species. Mol Biol Evol 20:424-434

REFERENCES

Goldfarb LG, Brown P, McCombie WR, Goldgaber D, Swergold GD, Wills PR, Cervenakova L, Baron H, Gibbs CJ, Jr., Gajdusek DC (1991) Transmissible familial Creutzfeldt-Jakob disease associated with five, seven, and eight extra octapeptide coding repeats in the PRNP gene. Proc Natl Acad Sci U S A 88:10926-10930

Goldfarb LG, Korczyn AD, Brown P, Chapman J, Gajdusek DC (1990a) Mutation in codon 200 of scrapie amyloid precursor gene linked to Creutzfeldt-Jakob disease in Sephardic Jews of Libyan and non-Libyan origin. Lancet 336:637-638

Goldfarb LG, Mitrova E, Brown P, Toh BK, Gajdusek DC (1990b) Mutation in codon 200 of scrapie amyloid protein gene in two clusters of Creutzfeldt-Jakob disease in Slovakia. Lancet 336:514-515

Goldfarb LG, Petersen RB, Tabaton M, Brown P, LeBlanc AC, Montagna P, Cortelli P, Julien J, Vital C, Pendelbury WW, et al. (1992) Fatal familial insomnia and familial Creutzfeldt-Jakob disease: disease phenotype determined by a DNA polymorphism. Science 258:806-808

Goldman JS, Miller BL, Safar J, de Tourreil S, Martindale JL, Prusiner SB, Geschwind MD (2004) When sporadic disease is not sporadic: the potential for genetic etiology. Arch Neurol 61:213-216

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725-736

Gonder MK, Oates JF, Disotell TR, Forstner MR, Morales JC, Melnick DJ (1997) A new west African chimpanzee subspecies? Nature 388:337

Goodman M (1999) The genomic record of Humankind's evolutionary roots. Am J Hum Genet 64:31-39

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9:585-598

Goodman S, Check E (2002) The great primate debate. Nature 417:684-687

Gordo I, Navarro A, Charlesworth B (2002) Muller's ratchet and the pattern of variation at a neutral locus. Genetics 161:835-848

Griffith JS (1967) Self-replication and scrapie. Nature 215:1043-1044

Harder A, Jendroska K, Kreuz F, Wirth T, Schafranka C, Karnatz N, Theallier-Janko A, Dreier J, Lohan K, Emmerich D, Cervos-Navarro J, Windl O, Kretzschmar HA, Nurnberg P, Witkowski R (1999) Novel twelve-generation kindred of fatal familial insomnia from germany representing the entire spectrum of disease expression. Am J Med Genet 87:311-316

Harper JD, Lansbury PT, Jr. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences

of the time-dependent solubility of amyloid proteins. Annu Rev Biochem 66:385-407

Harris DA (1999) Cellular biology of prion diseases. Clin Microbiol Rev 12:429-444

Harris EE, Hey J (2001) Human populations show reduced DNA sequence variation at the factor IX locus. Curr Biol 11:774-778

Hayakawa T, Satta Y, Gagneux P, Varki A, Takahata N (2001) Alu-mediated inactivation of the human CMP- N-acetylneuraminic acid hydroxylase gene. Proc Natl Acad Sci U S A 98:11399-11404

Head MW, Northcott V, Rennison K, Ritchie D, McCardle L, Bunn TJ, McLennan NF, Ironside JW, Tullo AB, Bonshek RE (2003) Prion protein accumulation in eyes of patients with sporadic and variant Creutzfeldt-Jakob disease. Invest Ophthalmol Vis Sci 44:342-346

Hitoshi S, Nagura H, Yamanouchi H, Kitamoto T (1993) Double mutations at codon 180 and codon 232 of the PRNP gene in an apparently sporadic case of Creutzfeldt-Jakob disease. J Neurol Sci 120:208-212

Holt TA, Phillips J (1988) Bovine spongiform encephalopathy. Br Med J (Clin Res Ed) 296:1581-1582

Hoque MZ, Kitamoto T, Furukawa H, Muramoto T, Tateishi J (1996) Mutation in the prion protein gene at codon 232 in Japanese patients with

Creutzfeldt-Jakob disease: a clinicopathological, immunohistochemical and transmission study. Acta Neuropathol (Berl) 92:441-446

Horiuchi M, Caughey B (1999) Prion protein interconversions and the transmissible spongiform encephalopathies. Structure Fold Des 7:R231-240

Hsiao K, Baker HF, Crow TJ, Poulter M, Owen F, Terwilliger JD, Westaway D, Ott J, Prusiner SB (1989) Linkage of a prion protein missense variant to Gerstmann-Straussler syndrome. Nature 338:342-345

Hsiao K, Meiner Z, Kahana E, Cass C, Kahana I, Avrahami D, Scarlato G, Abramsky O, Prusiner SB, Gabizon R (1991) Mutation of the prion protein in Libyan Jews with Creutzfeldt-Jakob disease. N Engl J Med 324:1091-1097

Hsiao KK, Scott M, Foster D, Groth DF, DeArmond SJ, Prusiner SB (1990) Spontaneous neurodegeneration in transgenic mice with mutant prion protein. Science 250:1587-1590

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153-159

Huillard d'Aignaux J, Costagliola D, Maccario J, Billette de Villemeur T, Brandel JP, Deslys JP, Hauw JJ, Chaussain JL, Agid Y, Dormont D, Alperovitch A (1999) Incubation period of Creutzfeldt-Jakob disease in human growth hormone recipients in France. Neurology 53:1197-1201

Huttley GA, Easteal S, Southey MC, Tesoriero A, Giles GG, McCredie MR, Hopper JL, Venter DJ (2000) Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Australian Breast Cancer Family Study. Nat Genet 25:410-413

Ironside JW, Bell JE (1997) Florid plaques and new variant Creutzfeldt-Jakob disease. Lancet 350:1475

Jeffrey M, Wells GA (1988) Spongiform encephalopathy in a nyala (Tragelaphus angasi). Vet Pathol 25:398-399

Johnson RT, Gibbs CJ, Jr. (1998) Creutzfeldt-Jakob disease and related transmissible spongiform encephalopathies. N Engl J Med 339:1994-2004

Kaessmann H, Heissig F, von Haeseler A, Paabo S (1999a) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat Genet 22:78-81

Kaessmann H, Wiebe V, Paabo S (1999b) Extensive nuclear DNA sequence diversity among chimpanzees. Science 286:1159-1162

Kaessmann H, Wiebe V, Weiss G, Paabo S (2001) Great ape DNA sequences reveal a reduced diversity and an expansion in humans. Nat Genet 27:155-156

Kahana E, Alter M, Braham J, Sofer D (1974) Creutzfeldt-jakob disease: focus among Libyan Jews in Israel. Science 183:90-91

Kahana E, Zilber N, Abraham M (1991) Do Creutzfeldt-Jakob disease patients of Jewish Libyan origin have unique clinical features? Neurology 41:1390-1392

Karaman MW, Houck ML, Chemnick LG, Nagpal S, Chawannakul D, Sudano D, Pike BL, Ho VV, Ryder OA, Hacia JG (2003) Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. Genome Res 13:1619-1630

Kascsak RJ, Rubenstein R, Carp RI (1991) Evidence for biological and structural diversity among scrapie strains. Curr Top Microbiol Immunol 172:139-152

Kayser M, Brauer S, Stoneking M (2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. Mol Biol Evol 20:893-900

Kidd KK, Pakstis AJ, Speed WC, Kidd JR (2004) Understanding human DNA sequence variation. J Hered 95:406-420

Kimberlin RH (1986) Scrapie: how much do we really understand? Neuropathol Appl Neurobiol 12:131-147

Kimberlin RH (1990) Bovine spongiform encephalopathy. Taking stock of the issues. Nature 345:763-764

Kirkwood JK, Wells GA, Wilesmith JW, Cunningham AA, Jackson SI (1990) Spongiform encephalopathy in an arabian oryx (Oryx leucoryx) and a greater kudu (Tragelaphus strepsiceros). Vet Rec 127:418-420

Kitamoto T, Ohta M, Doh-ura K, Hitoshi S, Terao Y, Tateishi J (1993) Novel missense variants of prion protein in Creutzfeldt-Jakob disease or Gerstmann-Straussler syndrome. Biochem Biophys Res Commun 191:709-714

Kocisko DA, Come JH, Priola SA, Chesebro B, Raymond GJ, Lansbury PT, Caughey B (1994) Cell-free formation of protease-resistant prion protein. Nature 370:471-474

Kong A, Barnard J, Gudbjartsson DF, Thorleifsson G, Jonsdottir G, Sigurdardottir S, Richardsson B, Jonsdottir J, Thorgeirsson T, Frigge ML, Lamb NE, Sherman S, Gulcher JR, Stefansson K (2004) Recombination rate and reproductive success in humans. Nat Genet 36:1203-1206

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241-247

Kovacs GG, Trabattoni G, Hainfellner JA, Ironside JW, Knight RS, Budka H (2002) Mutations of the prion protein gene phenotypic spectrum. J Neurol 249:1567-1582

Krakauer DC, Zanotto PM, Pagel M (1998) Prion's progress: patterns and rates of molecular evolution in relation to spongiform disease. J Mol Evol 47:133-145

Krawczak M, Ball EV, Cooper DN (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am J Hum Genet 63:474-488

Kreitman M, Di Rienzo A (2004) Balancing claims for balancing selection. Trends Genet 20:300-304

Laplanche JL, Hachimi KH, Durieux I, Thuillet P, Defebvre L, Delasnerie-Laupretre N, Peoc'h K, Foncin JF, Destee A (1999) Prominent psychiatric features and early onset in an inherited prion disease with a new insertional mutation in the prion protein gene. Brain 122 (Pt 12):2375-2386

Lee HS, Brown P, Cervenakova L, Garruto RM, Alpers MP, Gajdusek DC, Goldfarb LG (2001) Increased susceptibility to Kuru of carriers of the PRNP 129 methionine/methionine genotype. J Infect Dis 183:192-196

Lee IY, Westaway D, Smit AF, Wang K, Seto J, Chen L, Acharya C, Ankener M, Baskin D, Cooper C, Yao H, Prusiner SB, Hood LE (1998) Complete

genomic sequence and analysis of the prion protein gene region from three mammalian species. Genome Res 8:1022-1037

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175-195

Lugaresi E, Montagna P, Baruzzi A, Cortelli P, Tinuper P, Zucconi M, Gambetti PL, Medori R (1986) [Familial insomnia with a malignant course: a new thalamic disease]. Rev Neurol (Paris) 142:791-792

Mahal SP, Asante EA, Antoniou M, Collinge J (2001) Isolation and functional characterisation of the promoter region of the human prion protein gene. Gene 268:105-114

Manson JC, Jamieson E, Baybutt H, Tuzi NL, Barron R, McConnell I, Somerville R, Ironside J, Will R, Sy MS, Melton DW, Hope J, Bostock C (1999) A single amino acid alteration (101L) introduced into murine PrP dramatically alters incubation time of transmissible spongiform encephalopathy. Embo J 18:6855-6864

Marlar RA, Leonard BL, Billman BR, Lambert PM, Marlar JE (2000) Biochemical evidence of cannibalism at a prehistoric Puebloan site in southwestern Colorado. Nature 407:74-78

Martinez-Arias R, Calafell F, Mateu E, Comas D, Andres A, Bertranpetit J (2001) Sequence variability of a human pseudogene. Genome Res 11:1071-1085

Masters CL, Gajdusek DC, Gibbs CJ, Jr. (1981) The familial occurrence of Creutzfeldt-Jakob disease and Alzheimer's disease. Brain 104:535-558

Masters CL, Harris JO, Gajdusek DC, Gibbs CJ, Jr., Bernoulli C, Asher DM (1979) Creutzfeldt-Jakob disease: patterns of worldwide occurrence and the significance of familial and sporadic clustering. Ann Neurol 5:177-188

McCormack JE, Baybutt HN, Everington D, Will RG, Ironside JW, Manson JC (2002) PRNP contains both intronic and upstream regulatory regions that may influence susceptibility to Creutzfeldt-Jakob Disease. Gene 288:139-146

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654

Mead S, Mahal SP, Beck J, Campbell T, Farrall M, Fisher E, Collinge J (2001) Sporadic--but not variant--Creutzfeldt-Jakob disease is associated with polymorphisms upstream of PRNP exon 1. Am J Hum Genet 69:1225-1235

Mead S, Stumpf MP, Whitfield J, Beck JA, Poulter M, Campbell T, Uphill JB, Goldstein D, Alpers M, Fisher EM, Collinge J (2003) Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. Science 300:640-643

Meiner Z, Gabizon R, Prusiner SB (1997) Familial Creutzfeldt-Jakob disease. Codon 200 prion disease in Libyan Jews. Medicine (Baltimore) 76:227-237

Mestel R (1996) Putting prions to the test. Science 273:184-189

Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science 265:1193-1201

Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol 11:715-724

Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. Genetics 156:297-304

Navarro A, Barbadilla A, Ruiz A (2000) Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. Genetics 155:685-698

Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. Science 300:321-324

Neuhauser C, Krone SM (1997) The genealogy of samples in models with selection. Genetics 145:519-534

Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154:931-942

Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. Heredity 86:641-647

Nielsen R, Hubisz MJ, Clark AG (2004) Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. Genetics 168:2373-2382

Nielsen R, Signorovitch J (2003) Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. Theor Popul Biol 63:245-255

Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83-90

Olson MV, Varki A (2003) Sequencing the chimpanzee genome: insights into human evolution and disease. Nat Rev Genet 4:20-28

Palmer MS, Dryden AJ, Hughes JT, Collinge J (1991) Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease. Nature 352:340-342

Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ, Cohen FE, et al. (1993) Conversion of alpha-helices into

beta-sheets features in the formation of the scrapie prion proteins. Proc Natl Acad Sci U S A 90:10962-10966

Parchi P, Giese A, Capellari S, Brown P, Schulz-Schaeffer W, Windl O, Zerr I, Budka H, Kopp N, Piccardo P, Poser S, Rojiani A, Streichemberger N, Julien J, Vital C, Ghetti B, Gambetti P, Kretzschmar H (1999) Classification of sporadic Creutzfeldt-Jakob disease based on molecular and phenotypic analysis of 300 subjects. Ann Neurol 46:224-233

Prusiner SB (1991a) Molecular biology of prion diseases. Science 252:1515-1522

Prusiner SB (1991b) Molecular biology of prions causing infectious and genetic encephalopathies of humans as well as scrapie of sheep and BSE of cattle. Dev Biol Stand 75:55-74

Prusiner SB (1998) Prions. Proc Natl Acad Sci U S A 95:13363-13383

Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Paabo S (2005) Fine-scale recombination patterns differ between chimpanzees and humans. Nat Genet

Ptak SE, Roeder AD, Stephens M, Gilad Y, Paabo S, Przeworski M (2004) Absence of the TAP2 human recombination hotspot in chimpanzees. PLoS Biol 2:e155

Riek R, Wider G, Billeter M, Hornemann S, Glockshuber R, Wuthrich K (1998) Prion protein NMR structure and familial human spongiform encephalopathies. Proc Natl Acad Sci U S A 95:11667-11672

Rodriguez-Martinez AB, Barreau C, Coupry I, Yague J, Sanchez-Valle R, Galdos-Alcelay L, Ibanez A, Digon A, Fernandez-Manchola I, Goizet C, Castro A, Cuevas N, Alvarez-Alvarez M, de Pancorbo MM, Arveiler B, Zarranz JJ (2005) Ancestral origins of the prion protein gene D178N mutation in the Basque Country. Hum Genet

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552-569

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381-2385

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837

Saborio GP, Soto C, Kascsak RJ, Levy E, Kascsak R, Harris DA, Frangione B (1999) Cell-lysate conversion of prion protein into its protease-resistant

isoform suggests the participation of a cellular chaperone. Biochem Biophys Res Commun 258:470-475

Safar J, Wille H, Itri V, Groth D, Serban H, Torchia M, Cohen FE, Prusiner SB (1998) Eight prion strains have PrP(Sc) molecules with different conformations. Nat Med 4:1157-1165

Sanchez-Valle R, Nos C, Yague J, Graus F, Dominguez A, Saiz A (2004) Clinical and genetic features of human prion diseases in Catalonia: 1993-2002. Eur J Neurol 11:649-655

Scott MR, Will R, Ironside J, Nguyen HO, Tremblay P, DeArmond SJ, Prusiner SB (1999) Compelling transgenetic evidence for transmission of bovine spongiform encephalopathy prions to humans. Proc Natl Acad Sci U S A 96:15137-15142

Seabury CM, Honeycutt RL, Rooney AP, Halbert ND, Derr JN (2004) Prion protein gene (PRNP) variants and evidence for strong purifying selection in functionally important regions of bovine exon 3. Proc Natl Acad Sci U S A 101:15142-15147

Seitelberger F (1981) Straussler's disease. Acta Neuropathol Suppl (Berl) 7:341-343

Shibuya S, Higuchi J, Shin RW, Tateishi J, Kitamoto T (1998) Codon 219 Lys allele of PRNP is not found in sporadic Creutzfeldt-Jakob disease. Ann Neurol 43:826-828

Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141:413-429

Soldevila M, Calafell F, Andres AM, Yague J, Helgason A, Stefansson K, Bertranpetit J (2003) Prion susceptibility and protective alleles exhibit marked geographic differences. Hum Mutat 22:104-105

Soto C, Saborio GP (2001) Prions: disease propagation and disease therapy by conformational transmission. Trends Mol Med 7:109-114

Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci U S A 100:1056-1061

Stahl N, Baldwin MA, Teplow DB, Hood L, Gibson BW, Burlingame AL, Prusiner SB (1993) Structural studies of the scrapie prion protein using mass spectrometry and amino acid sequencing. Biochemistry 32:1991-2002

Stajich JE, Hahn MW (2005) Disentangling the effects of demography and selection in human history. Mol Biol Evol 22:63-73

Stedman HH, Kozyak BW, Nelson A, Thesier DM, Su LT, Low DW, Bridges CR, Shrager JB, Minugh-Purvis N, Mitchell MA (2004) Myosin gene mutation correlates with anatomical changes in the human lineage. Nature 428:415-418

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293:489-493

Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162-1169

Storz JF, Payseur BA, Nachman MW (2004) Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol Biol Evol 21:1800-1811

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

Tavare S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. Genetics 145:505-518

Taylor DM (1989) Bovine spongiform encephalopathy and human health. Vet Rec 125:413-415

Telling GC, Scott M, Mastrianni J, Gabizon R, Torchia M, Cohen FE, DeArmond SJ, Prusiner SB (1995) Prion propagation in mice expressing human and chimeric PrP transgenes implicates the interaction of cellular PrP with another protein. Cell 83:79-90

Thorgeirsdottir S, Sigurdarson S, Thorisson HM, Georgsson G, Palsdottir A (1999) PrP gene polymorphism and natural scrapie in Icelandic sheep. J Gen Virol 80 (Pt 9):2527-2534

Uddin M, Wildman DE, Liu G, Xu W, Johnson RM, Hof PR, Kapatos G, Grossman LI, Goodman M (2004) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. Proc Natl Acad Sci U S A 101:2957-2962

Vallender EJ, Lahn BT (2004) Positive selection on the human genome. Hum Mol Genet 13 Spec No 2:R245-254

Valleron AJ, Boelle PY, Will R, Cesbron JY (2001) Estimation of epidemic size and incubation time based on age characteristics of vCJD in the United Kingdom. Science 294:1726-1728

van Gool WA, Hensels GW, Hoogerwaard EM, Wiezer JH, Wesseling P, Bolhuis PA (1995) Hypokinesia and presenile dementia in a Dutch family with a novel insertion in the prion protein gene. Brain 118 (Pt 6):1565-1571

Varki A (2000) A chimpanzee genome project is a biomedical imperative. Genome Res 10:1065-1070

Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K (2001) The discovery of single-nucleotide polymorphisms--and inferences about human demographic history. Am J Hum Genet 69:1332-1347

Wall JD (2003) Estimating ancestral population sizes and divergence times. Genetics 163:395-404

Wall JD, Frisse LA, Hudson RR, Di Rienzo A (2003) Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. Am J Hum Genet 73:1330-1340

Watterson GA (1977) Heterosis or neutrality? Genetics 85:789-814

Wells GA, Scott AC, Johnson CT, Gunning RF, Hancock RD, Jeffrey M, Dawson M, Bradley R (1987) A novel progressive spongiform encephalopathy in cattle. Vet Rec 121:419-420

Will RG, Ironside JW, Zeidler M, Cousens SN, Estibeiro K, Alperovitch A, Poser S, Pocchiari M, Hofman A, Smith PG (1996) A new variant of Creutzfeldt-Jakob disease in the UK. Lancet 347:921-925

Willoughby K, Kelly DF, Lyon DG, Wells GA (1992) Spongiform encephalopathy in a captive puma (Felis concolor). Vet Rec 131:431-434

Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D (2005) Comparison of Fine-Scale Recombination Rates in Humans and Chimpanzees. Science

Windl O, Dempster M, Estibeiro JP, Lathe R, de Silva R, Esmonde T, Will R, Springbett A, Campbell TA, Sidle KC, Palmer MS, Collinge J (1996)

Genetic basis of Creutzfeldt-Jakob disease in the United Kingdom: a systematic analysis of predisposing mutations and allelic variation in the PRNP gene. Hum Genet 98:259-264

Winter H, Langbein L, Krawczak M, Cooper DN, Jave-Suarez LF, Rogers MA, Praetzel S, Heidt PJ, Schweizer J (2001) Human type I hair keratin pseudogene phihHaA has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the Pan-Homo divergence. Hum Genet 108:37-42

Wise CA, Sraml M, Rubinsztein DC, Easteal S (1997) Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. Mol Biol Evol 14:707-716

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution 15:496-503

Yi S, Ellsworth DL, Li WH (2002) Slow molecular clocks in Old World monkeys, apes, and humans. Mol Biol Evol 19:2191-2198

Young S, Slocombe RF (2003) Prion-associated spongiform encephalopathy in an imported Asiatic golden cat (Catopuma temmincki). Aust Vet J 81:295-296

Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li WH (2003) Low nucleotide diversity in chimpanzees and bonobos. Genetics 164:1511-1518

Yunis JJ, Prakash O (1982) The origin of man: a chromosomal pictorial legacy. Science 215:1525-1530

Zarranz JJ, Digon A, Atares B, Arteagoitia JM, Carrera N, Fernandez-Manchola I, Fernandez-Martinez M, Fernandez-Maiztegui C, Forcadas I, Galdos L, Ibanez A, Lezcano E, Marti-Masso JF, Mendibe MM, Urtasun M, Uterga JM, Saracibar N, Velasco F, Gonzalez de Galdeano L (2005) Familial prion diseases in the Basque Country (Spain). Neuroepidemiology 24:103-109

Zhang XM, Cathala G, Soua Z, Lefranc MP, Huck S (1996) The human T-cell receptor gamma variable pseudogene V10 is a distinctive marker of human speciation. Immunogenetics 43:196-203

Zilber N, Kahana E, Abraham M (1991) The Libyan Creutzfeldt-Jakob disease focus in Israel: an epidemiologic evaluation. Neurology 41:1385-1389

Books

Prion Diseases. Edited by John Collinge and Mark S. Palmer. Oxford University Press,1997 New York, U.S.A.

Unconventional viruses and the origin and disappearance of Kuru. Nobel lecture, december 13, 1976. Carleton D. Gajdusek. National Institutes of Health, Bethesda, Maryland, U.S.A.

Human Evolutionary Genetics. Origins, peoples and disease. Edited by M.A. Jobling, M.E. Hurles, C. Tyler-Smith. Garland Science, New York, U.S.A.

Tornarem a lluitar, tornarem a sofrir, tornarem a vèncer.

Lluís Companys.

No permitáis que ésto acabe así.
Contad que he dicho algo.

Darreres paraules de Pancho Villa