Drug discovery and computational evolutionary analysis

Joanna D. Holbrook and Philippe Sanseau

GlaxoSmithKline, Molecular Discovery Research, Bioinformatics Analysis, Stevenage SG1 2NY, United Kingdom

Drug discovery remains a difficult business with a very high level of attrition. Many steps in this long process use data generated from various species. One key challenge is to successfully translate the pre-clinical findings of target validation and safety studies in animal models to diverse human beings in the clinic. Advanced computational evolutionary analysis techniques combined with the increasing availability of sequence information enable the application of systematic evolutionary approaches to targets and pathways of interest to drug discovery. These analyses have the potential to increase our understanding of experimental differences observed between species.

The discovery of new drugs to treat human diseases is a long and difficult procedure. Despite increased investment in R&D the success rate in pharmaceutical pipelines remains low [1]. One reason for this attrition is the difficulty of successfully translating safety and efficacy studies performed in animal models to humans. A possible cause is that the biology of the drug target itself and/or the disease pathways targeted by small molecules differ across species and also between human individuals. These differences could be due to changes acquired during the evolution of the species and their genome. One way of analysing this is to determine the evolutionary history of the genes encoding the drug targets. For example, this could help to choose the appropriate animal model in pre-clinical studies to reflect human biology and thus to ultimately reduce the cost of target validation.

The development of high-throughput phylogenetic analysis tools and evolutionary analysis techniques such as selection pressure detection methods combined with the availability of genomic sequence data from multiple species is making evolutionary studies increasingly practical. Moreover, the availability of genetic information from different ethnic groups through projects such as HapMap [2] enables the study of gene evolution within human populations. Therefore, it is now possible to analyse evolution across relatively ancient (mammalian) and relatively recent (human) timescales in a genomic context and to an unprecedented extent. To illustrate the use of these approaches in a pharmacological context the term pharmacophylogenomics was coined [3].

Corresponding author: Sanseau, P. (philippe.x.sanseau@gsk.com)

Phylogenetic assignment of orthology and paralogy

The drug discovery process requires the use of different species (e.g. mouse, rat, dog) as pre-clinical animal models before a new compound is administered to humans. Pre-clinical experiments assume that the effect of the drug tested on the model species is comparable to that on humans, which can only be true if a functional equivalent of the human drug target exists in the experimental species. Hence, the importance to understand evolutionary relationships of drug targets for species of interest to drug discovery.

Orthology and paralogy are the first important concepts to introduce in evolutionary analysis. Briefly, orthologues are homologous genes that exist in different species as a result of speciation. Orthologues, commonly, have the same function and are often treated as comparable in closely related species. Paralogues are the results of gene duplication and are of interest, for example for the expansion of a tractable gene family or for specificity studies. Figure 1 illustrates the definitions of orthologues and paralogues. For more details, the reader should refer to a recent review [4]. It is important that orthology and paralogy analyses are performed on drug targets early during validation. A clear orthology relationship between a gene in an animal model and its equivalent in the human genome will support the use of that animal species in pre-clinical research. By contrast, the lack of an orthologue in a specific

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2007.08.015

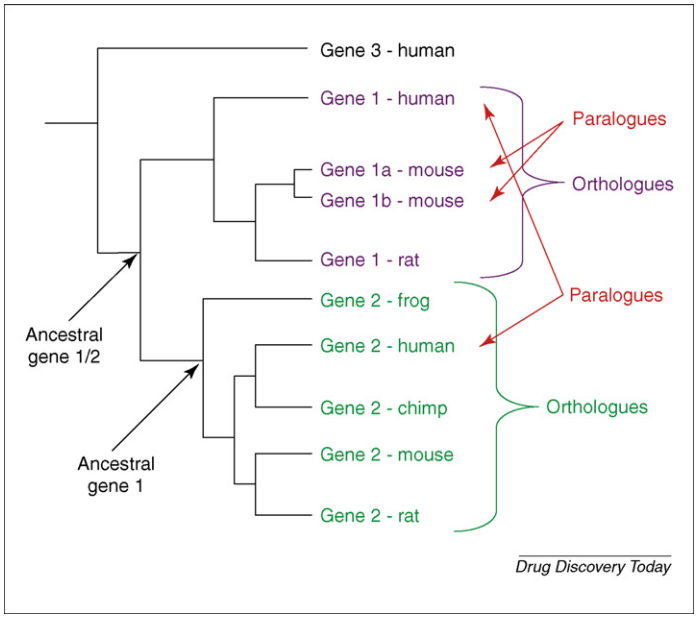

FIGURE 1

Definitions of the terms orthology and paralogy. *Orthologues*: Orthologues are products of speciation. They arise from a common ancestral gene. They are genes in different species that are more similar to each other than they are to any other gene in the same species. In phylogenetics, orthologues should cluster together into a clade and follow a taxonomic distribution within it. *Paralogues*: Paralogues are products of gene duplication. In the example above gene 1 and gene 2 duplicated from an ancestral gene and have then evolved into two orthologous clusters following speciation events. Mouse genes 1a and 1b have duplicated on the mouse lineage forming two paralogues that are together orthologous to human gene 1.

species would suggest the use of a different animal model. Paralogy also has significance for drug discovery. For example, closely paralogous genes may lead to off target effects with the compound tested binding to multiple targets in the same species. It is worth noting that evolutionary relationships for orthologues and paralogues are complicated by lineage-specific loss and duplication of genes. During evolution, an orthologue could be lost in one species while being retained in a sister species. Similarly, an ancestral gene could undergo duplication after a speciation event leading to two paralogous genes in one species and only one in its sister species.

A limitation to orthologue assignment is when the genomic sequence of a particular species is not complete. In that case, potential orthologues or paralogues could be missed or wrongly assigned. However, the availability of complete genomic sequences for many organisms has increased confidence in the reconstruction of evolutionary relationships. A simple analysis is to use a sequence similarity programme such as BLAST [5] or FASTA [6] and identify reciprocal top hits. However, this approach has limitations if the correct gene product has not been identified (e.g. splice variant), when a gene contains multiple protein domains or in cases of lineage-specific duplication or loss. Therefore, additional techniques are used such as the investigation of analogous chromosomal positions of genes between species known as synteny analysis or phylogenetic tree construction with inclusion of many homologous genes from multiple species. These techniques add a higher degree of confidence when combined. For a review of these methods please refer to reference [7].

After the evolutionary separation of a pair of orthologues, two genes may develop distinct functions in their respective organisms. Significant sequence divergence between orthologues is a signal for functional shifts. The topology of phylogenetic trees can be used to detect large divergence between orthologues indicating a functional shift. An example is the CYP2A family of cytochrome P450 enzymes where the rat liver isoform is in a different part of the tree than the human and mouse enzymes. This topological difference correlates with the non-identical mechanisms used by these species to metabolise the substrate coumarin: The rat metabolises the substrate to a toxic liver epoxide, whereas human and mouse do not [8].

Molecular selection analysis across species

A complementary approach to study evolution and identify potential functional shifts that could impact a drug target between species is to analyse molecular selective pressure at the coding sequence level. This can be measured by the ratio of the number of non-synonymous (amino acid changing) substitutions divided by the number of opportunities there are in that sequence for non-synonymous substitutions (d_N or K_a) and the number of synonymous (silent) substitutions divided by the number of opportunities there are in that sequence for synonymous substitutions (d_S or K_s) in the gene sequence. A normalised d_N/d_S rate ratio (known sometimes as the K_a/K_s ratio or ω) >1 is indicative of positive or adaptive selection also known as Darwinian selection [9]. This typically occurs when amino acid changing mutations lead to a selective advantage and are thus found more frequently in the evolved gene sequence than synonymous mutations. This is a hallmark of functional shift [10], since altered or novel functions induced by amino acid changes often provide selective advantage. Rate ratios of $d_N/d_S < 1$ indicate negative or purifying selection when amino acid changes are typically deleterious and so tend to be lost from the gene sequence during evolution. Neutral selection is shown by rate ratios of 1; this is common in non-coding regions of genes such as introns. Many methods exist to count these rate ratios; they vary by whether they correct for transition/transversion differences and unequal codon frequencies [11,12].

Early studies to detect positive selection pressure across species compared homologous sequences in a pairwise manner, averaging d_N and d_S over the whole coding sequence and across all the evolutionary time that has passed since the separation of the two homologues. This test is very conservative for two reasons:

1. It is unlikely that a majority of sites in the coding sequence will be under positive selection. The majority may be under negative selection with a subset of key residues under positive selection pressure. Therefore, the average d_N/d_S may not be >1 even when positive selection has acted on parts of the protein.
2. The positive selection pressure may not have been constantly acting on both genes since their separation; instead it may have acted on only one of them and maybe for a short time during their evolution.

Further developments of these methods have attempted to account for the two factors listed above (namely only a subset of residues evolving under positive selection pressure in a subset of the evolutionary time elapsed since the divergence of the

orthologues). In an interesting example, Choi and Lahn [13] used sliding window analysis to find areas of the coding sequence under positive selection pressure interspersed with areas under negative selection. Moreover, the two factors listed above are controlled for in the powerful methods implemented within the phylogenetic analysis by maximum likelihood (PAML) software package [14,15]:

- To address factor 1: PAML performs statistical tests to determine if a model of sequence evolution that allows a subset of residues (not necessarily contiguous) to evolve under positive selection fits the data (the known sequences) better than a neutral model of sequence evolution. This is known as sites analysis.
- To address factor 2: PAML performs statistical tests to determine if a model of sequence evolution that allows one branch of a phylogenetic tree to evolve under positive selection pressure while the other branches do not fit the data better than a neutral model of sequence evolution. This is known as branch analysis.
- To address both factors together: PAML performs statistical tests to determine if a model of sequence evolution that allows a subset of residues in one branch of a phylogenetic tree to evolve under positive selection pressure fits the data better than a neutral model of sequence evolution. This is known as branch/sites analysis.

PAML also has the advantage that once positive selection pressure has been detected those residues acting under its influence can be identified [16–19]. For a detailed description of these approaches and others see reference [9].

Molecular selection analysis within human population

With the completion of the human genome sequence and the availability of genetic variation data such as single nucleotide polymorphisms (SNPs) for several major human populations from the HapMap project or Perlegen Sciences [2,20] it is now possible to study evolution at shorter timescales and especially molecular selective pressure within the human lineage [21,22]. Human alleles under recent adaptive/positive selection will induce a distinctive pattern or signature of SNP allelic frequencies in the adjacent chromosomal regions that can be detected from the background distribution of genetic variation, which is generally assumed to be under neutral selection [23]. Several statistical methods can be used for the detection of these signatures or 'selective sweeps' to identify a rapid population-specific spread of a beneficial polymorphism. Among the most popular approaches are the Tajima's D [24], Fu and Li's D* [25], Fay and Wu's H [26], F_{ST} [27] and p_{excess} [28] tests. For a review of the statistical methods see reference [21]. It is important to note that detection of adaptive selection in any one lineage is challenging because of population demography such as expansions or bottlenecks, which affect all genes in the genome [29]. Investigating recent adaptive evolution in the human lineage is of interest since it has been argued that some complex diseases such as diabetes or obesity may be present in modern humans because of natural selection in response to ancient *stimuli* [30,31], and recent selection pressure may be responsible for the differential epidemiology and response to drugs we see across ethnic groups.

A growing number of genes and loci are being proposed to be under positive selection in the human lineage (for review see reference [21]). For example, a very recent selective sweep over the past ~7000 years was observed in the lactase gene to adapt to adult milk consumption [32]. The availability of human genetic variation data [2,20] on a large scale has enabled the systematic analysis of the human genome for recent positive selection. Several such studies have recently been published with some agreement on genes and regions that have been acted on by positive selection pressure including genes that function in drug metabolism [22,33–37].

Evolutionary analysis in single genes

There are numerous examples in the literature of lineage-specific loss/duplication and positive selection pressure across and within species during the evolution of therapeutically relevant genes. Some of these examples are listed below and may serve as a warning that observations made in pre-clinical species might not be confirmed in humans. However, they may also explain why some diseases differ between humans and other animals.

Therapeutically relevant examples of lineage-specific loss/duplication

HTR3

An example of lineage loss is the serotonin receptor 3 subunits *HTR3C*, *HTR3D* and *HTR3E* [38]. These subunits have recently been shown to be functionally relevant in the serotonergic system [39], and it has been demonstrated that the *HTR3C*, *D* and *E* genes exist in human and other species such as dog, rabbit and cow but are absent from rat and mouse [40]. Several drugs have been developed that regulate serotonin receptor 3 activity, for example to treat irritable bowel syndrome (IBS) [41]. Owing to the absence of the genes in rodents, however, it is impossible to study the therapeutic relevance of the subunit genes in rodent models.

MRGs

The *Mas-related gene* (*MRG*) family can be used to illustrate gene lineage duplication. *MRGs* are a family of G-protein-coupled receptors (GPCRs) expressed in nociceptive neurons. They are considered potential drug targets to treat pain [42]. Although in rodents, *MRGs* are subdivided into three families containing more than 10 highly duplicated genes, only 4 completed *MRG* gene sequences were identified in humans with 9 additional pseudogenes [43]. This makes orthologue assignments between rodents and humans extremely difficult and deprives researchers of a convenient animal model to progress our understanding of *MRG* biology. Interestingly, the MRG receptors also show signals of positive selection pressure: The sequences encoding the transmembrane and cytoplasmic protein domains have d_N/d_S ratios <1, while the sequences encoding the extracellular protein domains have ratios >1 [13]. Since the extracellular regions are thought to be involved in the binding of ligands, the positive selection probably affects novel receptor–ligand pairings or modifies affinities to the same ligand. As MRGs are implicated in nociception, the selection pressure may have been due to a diversification of the response to adverse *stimuli* and may indicate differences in nociception between humans and rodents.

Therapeutically relevant examples of positive selection pressure between species

Leptin

Genetic studies have shown that the leptin protein is associated with obesity in mice [44]. This finding generated a lot of excitement and stimulated the pursuit of leptin as a drug target in humans. Benner *et al*. analysed the ratio of silent synonymous and non-synonymous substitutions in nucleotide sequences from leptin orthologues [45]. The authors found a high ratio of non-synonymous to synonymous changes in the branches leading to primates after the divergence with rodents. This indicates adaptive selection in primates since the divergence with rodents and may suggest a change of function between human and rodents, therefore calling into question the role of leptin in human obesity.

BRAC1

BRAC1 mutations increase susceptibility to breast cancer in women [46]. A phylogeny-based analysis suggested positive selection in the human and chimpanzee lineages [47]. In addition, amino acid changes between these two lineages and other primates were located in the RAD51 interaction domain of the gene suggesting a potential link with cancer. This observation raises questions about the aetiology of breast cancer within primates. Are there functional differences between human and chimpanzee *BRAC1* compared with other primates, as is suggested by the evidence for positive selection pressure having acted on these lineages?

PDYN

Tests to detect the action of positive selection have also been applied to non-coding regions of genes, such as regulatory sequences. In an analysis combining a computational approach and experimental assays it was shown that positive selection affected the *cis*-regulation of the human opioid neuropeptide precursor gene *prodynorphin* (*PDYN*) [48]. Interestingly, opoid neuropeptides are endogenous ligands of opiate receptors known to be implicated in pain transmission or behavioural processes [49,50]. Perhaps the positive selection seen in the regulation of human *PDYN* is indicative of adaptation in the human lineage.

Therapeutically relevant examples of positive selection pressure within human evolution

Capsase 12

The *Capsase* 12 gene exists in two forms: full length active (ancestral) and inactive truncated (derived) by a stop codon [51]. Individuals with the short form are more resistant to severe sepsis than those expressing the long form. It was shown that individuals with two copies of the inactive shorter form of the gene are ~7.8-fold more likely to avoid severe sepsis and to survive if they are affected [52]. The inactive shorter form has been reported at a frequency of about 80% in Africa. Molecular evolution analysis in different populations demonstrated recent positive selection during the past 100 000 years [51]. It suggests that the inactive form originated in Africa and was initially neutral or almost neutral, but with a growing human population (i.e. more sepsis) it acquired selective advantage, hence the spread of the inactive form.

DRD4

Human dopamine receptor D4 (*DRD4*) has been associated with behavioural disorders and is the target of several antipsychotic compounds [53]. The *DRD4* gene is highly polymorphic with most of its variation located in the region encoding the third intracellular loop that binds to G proteins and contains a 48-bp tandem repeat [54]. In that region, variant human alleles ranging from 2 (2R) to 11 (11R) repeats can be detected. An evolutionary analysis suggested that the allele with seven repeats (7R) started as a rare mutation before the upper Paleolithic era and increased to high frequency by positive selection [55,56].

Evolutionary analysis of groups of genes

As positive selection pressure causing functional shift in a single gene can be problematic for drug discovery, it would be useful to identify groups of genes and gene pathways that have been subject to positive selection pressure before target identification. The most efficient way of doing this is to scan the whole genome for positive selection pressure signals. As techniques become more sophisticated, it may be possible to correlate the evolutionary characteristics of a group of genes to known biomedical differences between species and hence discover functional relationships between genes and disease mechanisms.

Genome scans

Over the past few years, several groups have been looking at evolutionary patterns across large groups of genes to explore the relationship between disease and selection pressure. Smith and Eyre-Walker found higher d_N/d_S ratios ($P < 0.001$) between human and rodent orthologues were detected for disease genes (defined by the Online Mendelian Inheritance in Man (OMIM) resource that catalogues causative genes in inherited genetic disorders) than non-disease genes suggesting these are under weaker purifying selection or maybe more often subject to positive selection [57]. Huang *et al.* also compared human genes with their rodent orthologues, but in this case found d_N/d_S ratios for 844 disease genes were only modestly higher than those of non-disease genes ($P = 0.035$) [58].

With the completion of the chimpanzee genome sequence, some researchers have been interested in analysing selection between primates and how this relates to disease. Three studies found that genes positively selected in human and chimpanzee lineages are overrepresented in OMIM [59–61]. These findings are particularly interesting as the methodologies used by these studies differ. Contradicting these findings, the Chimpanzee sequencing consortium found no significant difference in the d_N/d_S ratios of disease genes in the human and chimpanzee lineages [62].

Positive selection of disease genes may be due to adaptive changes in response to the environment of early hominids that are maladaptive in the different conditions we live in today [63]. Another possibility is that highly derived characteristics that have evolved since the separation of chimpanzees and humans cause higher rates of disease than characteristics that have been subject to purifying selection for millions of years, for example are humans more susceptible to psychiatric disease than other animals because of our specialisation for higher cognitive function?

Several recent studies of selection pressure across the genome have attempted to identify molecular functions that are enriched for positively selected genes using the public domain gene ontology

TABLE 1

Summary of functional categories identified as being enriched for genes positively selected in the human lineage, in seven publications

Functional class	*Bustamante et al.* [60]	Nielsen *et al.* [66]	Rhesus macaque Genome sequencing and analysis consortium [67]	Clark *et al.* [59]	Bakewell *et al.* [61]	Arbiza *et al.* [68]	Chimpanzee sequencing and analysis consortium [62]
Defence/immunity	X	X	X			X	X
Signal transduction	X		X	X	X	X	
Reproduction	X	X	X				X
Apoptosis	X	X	X				X
Nucleotide metabolism	X			X		X	
Sensory perception	X	X	X	X		X	X
Transcription	X		X			X	
Subcellular transport			X	X	X	X	
Cellular structure			X	X		X	
Metabolism				X	X	X	X
Development				X	X		

Functional categories enriched for genes positively selected in human lineage. Categories are defined by both the GOA and PANTHER ontologies and grouped into larger classes. Confidence values are supplied by the publications, but as the studies are not directly comparable, the data have been simplified to X denoting a category that was reported to be significantly enriched for positively selected genes.

annotation (GOA) [64] or the Protein ANalysis THrough Evolutionary Relationships (PANTHER) ontology [65]. The classes highlighted by seven recent reports are summarised in Table 1 [59–62,66–68]. Although all the studies used slightly different comparisons and methodologies, there is some consensus regarding functional categories identified as enriched for positively selected genes in human evolution: defence/immunity, signal transduction, reproduction, apoptosis, nucleotide metabolism, sensory perception, transcription, subcellular transport, cellular structure, metabolism and development. Two additional points should also be considered:

1. PAML and other methods used in these studies are reported to be conservative in detecting positive selection [9] because of the two confounding factors described in the section entitled 'Molecular Selection Analysis Across Species'. Therefore, they would be expected to generate false negatives, especially given the relatively low numbers of taxa sampled (mostly two to three species were studied).
2. Genome scale automated orthologue detection and alignment is error prone and we have observed that including errors of this type in the starting data for PAML tends to give false positive results. It seems likely that these types of errors have been included in these studies and so we expect some false positives.

Despite these problems, the consensus seen here does suggest some interesting avenues for research. For instance, from a disease perspective it is worth noting the enrichment in genes involved in apoptosis (Table 1) suggests a potential link between oncogenesis and recent evolutionary changes in humans [69].

It is also tempting to speculate that characteristics of common psychiatric and neurological disorders in humans could be explained by positive selection acting on the hominid lineage. Huang *et al*. found that genes involved in neurological disease are well conserved between primates and tend to exhibit low d_N/d_S ratios [58]. However, as this analysis was performed using pairwise counting methods of d_N/d_S estimation across the entire length of the coding sequence, it is possible that any positive selection on a subset of residues within the gene would not be detected as the signal would be swamped by the purifying selection acting on the majority of residues in these crucial genes. In another study, 214 genes broadly covering nervous system biology were analysed [70]. On average, these genes have substantially higher d_N/d_S ratios in primates than rodents ($P < 0.0001$) suggesting adaptive evolution. By subclassification of these genes it was shown that genes involved in nervous system development have a greater d_N/d_S disparity between primates (human and macaques) then rodents (mice and rats) than housekeeping genes. Some of these genes code for chemically tractable cellular receptors (e.g. *DRD2*, *ORPM1*, *GRIK4*, *GRIN2A*, *CHRM5*) while others are associated with interesting phenotypes, such as defective social behaviour (*DVL1*) or changes in anxiety states (*ADCYAP1*). If neuropsychiatric disease genes have been subject to positive selection in recent human evolution it would indicate that translating findings from animal models to human in psychiatric diseases, such as schizophrenia, could be challenging, however it may also offer a new way of discovering genes implicated in neuropsychiatric function [71].

Conclusion

A better understanding of the evolutionary history of our drug targets will both minimise risk in drug development due to non-predictability of animal models and heterogeneity of human populations and concurrently maximise the advantages comparative genomics can bring to disease understanding. The broad relevance of the study of evolution to medicine with the need for more formal training has recently been highlighted [72,73].

The availability of the complete genome sequences of many species, human variation data and improving computational methods are making the evolutionary study of genes an interesting scientific endeavour. Genes, including those of interest to pharma, have been acted on by many different evolutionary pressures [3], and understanding their pattern of evolution could help drug

discovery, for example in the selection of the best disease models or aid disease understanding. Genome-wide studies are starting to confirm patterns of increased adaptive selection for some therapeutically relevant categories of genes. However, no studies have yet indicated if, for example targets of existing drugs have a defining evolutionary signature. In the human lineage, the influence of ethnicity on the efficacy and safety of medicines has been already established [74]. Considering this, it may be advantageous to start monitoring subjects from different ethnic groups in clinical trails for safety and differential response to new medicines. Tests to detect selective sweep events may be particularly pertinent in view of the continuing trend toward conducting clinical trials in populations from developing countries [75]. We suggest that a better evolutionary understanding of the potential issues that may arise during the testing of drugs in different populations will both minimise risk and maximise benefits to these populations by improving the chances of developing drugs that are efficacious in the populations being used to test them.

Acknowledgement

We would like to thank David Searls for valuable comments on the manuscript.

References

1 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–715
2 International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437, 1299–1320
3 Searls, D.B. (2003) Pharmacophylogenomics: genes, evolution and drug targets. *Nat. Rev. Drug Discov.* 2, 623
4 Koonin, E.V. (2005) Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* 39, 309–338
5 Stephen, F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
6 Pearson, W.R. and Lipman, D.J. (1998) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* 85, 2444–2448
7 McInerney, J.O. *et al.* (2006) Gene evolution and drug discovery. *Methods Mol. Biol.* 316, 87–109
8 Lake, B.G. (1999) Coumarin metabolism, toxicity and carcinogenicity: relevance for human risk assessment. *Food Chem. Toxicol.* 37, 423–453
9 Yang, Z. (2006). In Harvey P.H. and Robert, M.M. (eds), *Computational Molecular Evolution Oxford series in Ecology and Evolution*
10 Yang, Z. (2002) Inference of selection from multiple sequence alignments. *Curr. Opin. Genet. Dev.* 12, 688–694
11 Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426
12 Yang, Z. and Nielsen, R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17, 32–43
13 Choi, S.S. and Lahn, B.T. (2003) Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. *Gen. Res.* 13, 2252–2259
14 Yang, Z. (1998) Likelihood ratios tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15, 568–573
15 Nielsen, R. and Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936
16 Yang, Z. *et al.* (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449
17 Yang, Z. and Nielsen, R. (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917
18 Yang, Z. and Swanson, W.J. (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* 19, 49–57
19 Zhang, J. *et al.* (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479
20 Hinds, D.A. *et al.* (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* 307, 1072–1079
21 Sabeti, P.C. *et al.* (2006) Positive natural selection in the human lineage. *Science* 312, 1614–1620
22 Voight, B.F. *et al.* (2006) A map of recent positive selection in the human genome. *PloS. Biol.* 4, e72
23 Hellman, I. *et al.* (2003) A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* 72, 1527–1535
24 Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595
25 Fu, Y.X. and Li, W.H. (1993) Statistical tests of neutrality of mutations. *Genetics* 133, 693–709
26 Fay, J.C. and Wu, C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413
27 Akey, J.M. *et al.* (2002) Interrogating a high-density snp map for signatures of natural selection. *Genome Res.* 12, 1805–1814
28 Lewontin, R.C. and Krakauer, J. (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74, 175–195
29 Kreitman, M. (2000) Methods to detect selection in populations with applications to the human. *Annu. Rev. Genomics Hum. Genet.* 1, 539–559
30 Tishkoff, S.A. and Verelli, B.C. (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genomics Hum. Genet.* 4, 240–293
31 Di Rienzo, A. and Hudson, R.R. (2005) An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet.* 21, 596–601
32 Tishkoff, S.A. *et al.* (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39, 31–40
33 Carlson, C.S. *et al.* (2005) Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* 15, 1553–1565
34 Nielsen, R. *et al.* (2005) Genomic scans for selective sweep using SNP data. *Genome Res.* 15, 1666–1675
35 Kelley, J.L. *et al.* (2006) Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res.* 16, 980–989
36 Wang, E.T. *et al.* (2006) Global landscape of recent inferred Darwinian selection in *Homo sapiens*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 135–140
37 Tang, K. *et al.* (2007) A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 7, e171
38 Nielser, B. *et al.* (2003) Cloning, physical mapping and expression analysis of the human 5-HT_3 serotonin receptor-like genes *HTR3C*, *HTR3D* and *HTR3E*. *Gene* 310, 101–111
39 Nielser, B. *et al.* (2007) Characterization of the novel human serotonin receptor subunits 5-HT3C, 5-HT3D, and 5-HT3E. *Mol. Pharmacol.* 72, 8–17
40 Karnovsky, A.M. *et al.* (2003) A cluster of novel serotonin receptor 3-like genes on human chromosome 3. *Gene* 319, 137–148
41 Humphreys, P.P.A. *et al.* (1999) The therapeutic potential of 5-HT_3 receptor antagonists in the treatment of irritable bowel syndrome. *Aliment. Pharmacol. Ther.* 13, 31–38
42 Kinloch, R.A. and Cox, P.J. (2005) New targets for neuropathic pain therapeutics. *Expert Opin. Ther. Targets* 9, 685–698
43 Dong, X. *et al.* (2001) A diverse family of GPCRs expressed in specific subsets of nociceptive sensory neurons. *Cell* 106, 619–632
44 Chen, H. *et al.* (1996) Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in *db*/*db* mice. *Cell* 84, 491–495
45 Benner, S.A. *et al.* (2000) Functional inferences from reconstructed evolutionary biology involving rectified databases: an evolutionary grounded approach to functional genomics. *Res. Microbiol.* 151, 97–106
46 Miki, Y. *et al.* (1994) A strong candidate for the breast and ovarian cancer susceptibility gene *BRAC1*. *Science* 266, 66–71
47 Huttley, G.A. *et al.* (2000) Adaptive evolution of the tumour suppressor *BRAC1* in humans and chimpanzees. *Nat. Genet.* 25, 410–413
48 Rockman, M. *et al.* (2005) Ancient and recent positive selection transformed opioid *cis*-regulation in humans. *PloS Biol.* 3, e387
49 Cheng, H.Y. *et al.* (2002) DREAM is a critical transcriptional repressor for pain modulation. *Cell* 108, 31–43
50 Moles, A. *et al.* (2004) Deficit in attachment behaviour in mice lacking the mu-opioid receptor gene. *Science* 304, 1983–1986

51 Saleh, M. *et al.* (2004) Differential modulation of endotoxin responsiveness by human capsase-12 polymorphisms. *Nature* 429, 75–79
52 Xue, Y. *et al.* (2006) Spread of an inactive form of Capsase-12 in humans is due to recent positive selection. *Am. J. Hum. Genet.* 78, 659–670
53 Faranone, S.V. *et al.* (2001) Meta-analysis of the association between the 7-repeat allele of the dopamine D4 receptor gene and attention deficit hyperactivity disorder. *Am. J. Psychiatry* 158, 1052–1057
54 Jovanovic, V. *et al.* (1999) Comparative pharmacological and functional analysis of the human dopamine D4.2 and D4.10 receptor variants. *Pharmacogenetics* 9, 561–568
55 Ding, Y-C. *et al.* (2002) Evidence of positive selection acting at the human dopamine receptor D4 locus. *Proc. Natl. Acad. Sci. U.S.A.* 99, 309–314
56 Wang, E. *et al.* (2004) The genetic architecture of selection at the human dopamine receptor D4 (*DRD4*) gene locus. *Am. J. Hum. Genet.* 74, 931–944
57 Smith, N.G.C. and Eyre-Walker, A. (2003) Human disease genes: patterns and predictions. *Gene* 318, 169–175
58 Huang, H. *et al.* (2004) Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biol.* 5, R47
59 Clark, A.G. *et al.* (2003) Inferring non-neutral evolution from human-chimp-mouse orthologous gene trios. *Science* 302, 1960–1963
60 Bustamante, C.D. *et al.* (2005) Natural selection on protein-coding genes in the human genome. *Nature* 437, 1153–1157
61 Bakewell, M. *et al.* (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7489–7494
62 The Chimpanzee Sequencing and Analysis Consortium, (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69–87
63 Young, J.H. *et al.* (2005) Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. *PLoS Genet.* 1, e82
64 Camon, E. *et al.* (2004) The gene ontology annotation (GOA) database: sharing knowledge in uniprot with gene ontology. *Nucleic Acids Res.* 1, 32 (Database issue), D262–D266
65 Mi, H. *et al.* (2005) The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* 1, 33 (Database issue), D284–D288
66 Nielsen, R. *et al.* (2005) A scan for positively selected genes in the genomes of human and chimpanzees. *PloS Biol.* 3, e170
67 Rhesus Macaque Genome Sequencing and Anlaysis Consortium, (2007) Evolutionary and biomedical insights from the Rhesus Macaque Genome. *Science* 316, 222–234
68 Arbiza, L. *et al.* (2006) Positive selection, relaxation and acceleration in the evolution of the human and chimp genome. *PLoS Comput. Biol.* 2, 288–300
69 Crespi, B.J. and Summers, K. (2006) Positive selection in the evolution of cancer. *Biol. Rev.* 81, 407–424
70 Dorus, S. *et al.* (2004) Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* 119, 1027–1040
71 Crespi, B.J. (2006) The natural selection of psychosis. *Behav. Brain Sci.* 29, 410–411
72 Nesse, R.M. *et al.* (2006) Medicine needs evolution. *Science* 311, 1071
73 MacCallum, C.J. (2007) Does medicine without evolution make sense? *PLoS Biol.* 5, e112
74 Tate, S.K. and Goldstein, D.B. (2004) Will tomorrow's medicines work for everyone? *Nat. Genet.* 36 (Suppl. 11), S34–S42
75 Varmus, H. and Satcher, D. (1997) Ethical complexities of conducting research in developing countries. *N. Engl. J. Med.* 337, 1003–1005