In revision for *Nature Genetics*

(submitted October 18th, 2004)

An adaptive radiation model for the origin of new gene functions.

M. Pilar Francino

Evolutionary Genomics Department, DOE Joint Genome Institute

and Genomics Division, Lawrence Berkeley National Laboratory.

Correspondence should be addressed to: 2800 Mitchell Dr, Walnut Creek, CA, USA.

E-mail: mpfrancino@lbl.gov, Telephone: (925) 296 5872, Fax: (925) 296 5666.

The evolution of new gene functions is one of the keys to evolutionary innovation. Most novel functions result from gene duplication followed by divergence. However, the models hitherto proposed to account for this process are not fully satisfactory. The classic model of neofunctionalization[1] holds that the two paralogous gene copies resulting from a duplication are functionally redundant, such that one of them can evolve under no functional constraints and occasionally acquire a new function. This model lacks a convincing mechanism for the new gene copies to increase in frequency in the population and survive the mutational load expected to accumulate under neutrality, before the acquisition of the rare beneficial mutations that would confer new functionality. The subfunctionalization model[2] has been proposed as an alternative way to generate genes with altered functions. This model also assumes that new paralogous gene copies are functionally redundant and therefore neutral, but it predicts that relaxed selection will affect both gene copies such that some of the capabilities of the parent gene will disappear in one of the copies and be retained in the other. Thus, the functions originally present in a single gene will be partitioned between the two descendant copies. However, although this model can explain increases in gene number, it does not really address the main evolutionary question, which is the development of new biochemical capabilities. Recently, a new concept has been introduced into the gene evolution literature which is most likely to help solve this dilemma[3-6]. The key point is to allow for a period of natural selection for the duplication per se, before new function evolves, rather than considering gene duplication to be neutral as in the previous models. Here, I suggest a new model that draws on the advantage of postulating selection for gene

duplication, and proposes that bursts of adaptive gene amplification in response to specific selection pressures provide the raw material for the evolution of new function.

The amplification-mutagenesis phenomenon

A recent discovery in bacterial genetics catches the essence of this model in the act. This experimental finding has already solved another famous puzzle, the so-called adaptive mutagenesis phenomenon. For years, bacterial geneticists and evolutionary biologists were confronted by a seemingly unresolvable conflict: the possibility that some mutations could occur in a directed or "Lamarckian" fashion, since they were recovered at frequencies that exceeded expectation when bacteria were plated on a medium where these mutations would be advantageous[7]. Hendrickson *et al.* have solved this conflict by showing that the increase in mutation frequency is the direct consequence of an increase in the copy number of the target genes[3]. In the particular system they studied, bacteria harboring a defective, but somewhat leaky, *lacZ*$^-$ allele produced elevated numbers of *lacZ*$^+$ revertants when plated on lactose minimal medium. Most of these mutants appeared, not during growth in liquid before plating, but after a period of very slow growth on the lactose plates. Interestingly, the first step in this process was a substantial amplification of the *lacZ*$^-$ allele, which allowed bacteria to survive and reproduce on lactose by producing very large amounts of the defective enzyme encoded by their *lacZ*$^-$ allele. This gene expansion was selectively advantageous, since it amplified the minimal activity of the *lacZ*$^-$ allele to a level that permitted cell survival. The presence of an elevated number of copies of the *lacZ*$^-$ allele multiplied the likelihood of occurrence of a mutation that would restore LacZ$^+$ activity and full growth potential on lactose. Once a

gene copy acquired the reverting mutation and selectively spread through the bacterial population, the rest of the gene copies, now superfluous, rapidly disappeared.

Other cases of adaptation by gene amplification

Adaptation by means of gene duplications and larger amplifications has been documented in many other cases in bacteria[8-11], yeast[12], insects[13] and mammalian cell lines[14]. In bacteria, gene amplification has been demonstrated to underlie some instances of resistance to antibiotics and heavy metals[8], as well as experimental adaptation to growth at high temperature[9], and on limiting[8,10] or unusual[8,11] carbon sources. Adaptation to growth in glucose-limited chemostats through gene amplification has also been reported in yeast[12]. In mammalian cell cultures selected for antibody production, high productivity is attained by subclones in which light and heavy chain genes are amplified as extended arrays on diverse chromosomes[14]. In both laboratory strains and natural populations of the mosquito *Culex pipiens*, insecticide resistance is achieved by an overproduction of esterases due to gene amplification[13]. The widespread use of gene amplification by all kinds of organisms as an adaptive response to artificial selection suggests that this phenomenon is likely to be equally relevant in nature. In fact, gene amplifications have been implicated in enhanced virulence for some bacterial pathogens, as well as in increased production or fixation of host-required nutrients in symbionts[8]. The evolution of new gene functions in the wild, both in prokaryotes and in eukaryotes, may follow gene amplification[8] through a mechanism analogous to that described by Hendrickson *et al*[3,15].

The adaptive radiation model for the origin of new gene functions

Here, I propose a new model for the origin of new gene functions, an adaptive radiation model, which integrates the amplification-mutagenesis phenomenon observed in bacterial genetics with the patterns encountered in nature during the radiation of organismal lineages into new ecological niches. The adaptive radiation model postulates that new gene functions evolve in rapid, punctuated bursts when new biochemical niches appear, through large, selected amplifications of the best preadapted genes, followed by competition among the gene copies present throughout the population for the occupancy of the niche. When the fitness of an organism would benefit from a new molecular function, the newly opened biochemical niche may initially be filled by a suboptimal protein with some level of preadaptation to the novel function. Likely scenarios might be, for instance, the recognition of a novel compound in the environment or the establishment of a new regulatory interaction, where the preadapted proteins could be a membrane receptor currently recognizing a similar chemical or a transcription factor recognizing a similar DNA sequence or interacting protein. Presumably, the affinity of the preadapted protein for the new target molecule would be low, so that an increase in the amount of protein produced would likely be advantageous and the frequency of organisms carrying a gene amplification would increase by positive selection (assuming the benefits of binding the new target molecule offset the deleterious consequences of binding an excess of the original one). As in the *lacZ* system, gene amplification could initially provide the means to attain biologically significant levels of protein functionality, as well as the substrate for the evolution of improved function (**Fig. 1**).

The capacity of proteins to expand their substrate ranges has been well documented during experimental evolution[16,17]. Enzymes subject to directed laboratory selection under high rates of mutation and recombination rapidly increase their level of activity on new substrates by several orders of magnitude over the wild type[17]. *In vivo*, amplification of target genes may be the easiest and safest way to increase the likelihood of acquiring and bringing together adaptive mutations, rather than generally increasing the mutation and recombination rates at the genome-wide level. In addition, although new functions can evolve without affecting the initial activity of a protein, so that gene duplication is not a formal prerequisite for their evolution[17], gene amplification will multiply the possibilities of adaptation by freeing gene copies from their original constraints and providing alternate pathways for evolutionary tinkering.

Exploration of a new niche may be initiated by any kind of mechanism that increases the copy number of a preadapted gene. Genomes are more dynamic and plastic than previously realized, with many kinds of duplications and other rearrangements being common. Tandem, segmental and whole genome duplications have all been documented as sources of duplicates with new functions. Nevertheless, duplicative processes involving large portions of the genome are likely to produce fewer copies of the duplicate region. Further amplification within large duplicate regions of the specific genes with preadaptations for useful new functions could thus facilitate the process of adaptation.

Advantages of the adaptive radiation model

This proposal circumvents the principal hurdles of the neofunctionalization model by postulating:

a) a selective advantage for gene amplification per se, *followed* by the evolution of new gene function,

b) a multiplication of the probability of acquisition of beneficial mutations due to increased target sites,

c) the possibility of having different gene copies exploring different zones of the adaptive landscape, including fitness valleys for some of them,

d) the possibility of recombination among gene copies to bring together different beneficial mutations, and

e) the possibility of sequential acquisitions of beneficial mutations alternating with rounds of amplification of the best adapted gene copies at every step.

Predictions of the adaptive radiation model

This adaptive radiation model for the evolution of new gene functions makes predictions that are readily testable in our days of extensive genomic data for diverse species. Notably, the three main general predictions of this model are:

1) evolution of new functions after punctuated bursts of gene amplification and paralog fixation in response to specific selection pressures,

2) an initial period of purifying or positive selection on the paralogous gene copies, including those that eventually become pseudogenes, and

3) generation of numerous pseudogenes and eventual pseudogene loss accompanying the successful establishment of a new gene function or group of related functions.

Evidence for the adaptive radiation model

I suggest that the patterns predicted by the adaptive radiation model are detected in the current literature on gene family and genome evolution in both prokaryotes and eukaryotes, and moreover, that the adaptive radiation model explains certain observations that could not be accounted for by previous evolutionary models.

1) Punctuated bursts of gene amplification and paralog fixation in response to specific selection pressures

The association between gene duplication and selective episodes is indicated by the fact that duplications are most prevalent for functional classes otherwise known to be subject to frequent episodes of positive selection. Across bacteria, archaea and eukaryotes, extensive genome-wide analyses have revealed that gene duplications are most common in functional classes involved in species-level adaptations, or even in intraspecific adaptations to niche variation, such as transcription factors, environmental response regulators, transport proteins, efflux pumps and other cell surface and secreted proteins[4,18]. Interestingly, it has also been shown that, in bacteria, genes acquired by horizontal transfer undergo more duplication events than the ancestral, vertically transmitted genes of the same genome[19]. Horizontally transferred genes are most likely to undergo natural selection in order to adapt to the genomic environment and to the specific functional requirements of their new host. In a similar manner, species-specific genes in yeasts, which have no homologues in current databases, often form paralogous

families, indicating rapid expansion after their recent origin[20]. These genes, which are likely to encode new functions and to have appeared either by horizontal transfer from an unknown source or by novel combinations and modifications of coding and/or noncoding yeast DNA, are also most likely to have been subject to recent or ongoing selection in order to develop their new function or to adapt to their new host genome.

Under the adaptive radiation model, many of the largest gene families detected in all genomes would be transitory stages, present during limited periods of time during which a novel biochemical niche is occupied. This is consistent with the fact that the largest gene expansions are mostly confined to shallow phylogenetic depths. Indeed, the expansion coefficient, which is the ratio of the number of genes in a lineage-specific cluster to the total membership of that class of paralogs in a given genome[21], is remarkably biased towards values close to one, indicative of lineage-specific bursts within otherwise slowly proliferating gene families. Moreover, the expansion coefficient increases at shallower phylogenetic depths, such that more narrowly defined lineages contain a greater excess of specific expansions. This most likely reflects the short existence span of large gene expansions, rather than a recent acceleration of gene duplication across all types of organisms. Lineage-specific gene expansions have been identified within several sets of closely related species, including within mycoplasmas, spirochaetes, gamma-proteobacteria, epsilon-proteobacteria and methanogens[22], yeasts[20,21], as well as within humans and great apes[23]. Variation in the pattern of paralogous copy loss and retention has been documented even at the human population level[24,25], attesting to the recent origin and evolution of some amplified gene families. In fact, copy number polymorphisms contribute significantly to the observed genomic

variation between humans and are particularly abundant for genes involved in neurodevelopment[26].

The distribution of gene family sizes, for both genes and pseudogenes, in viral, bacterial, archaeal and eukaryotic genomes, is also consistent with the adaptive radiation model. Several studies have documented that the distribution of gene family sizes can be approximated by power laws[27-31]. Power law distributions feature very long tails, in this case indicating an overrepresentation of gene families that are much larger than the rest. Notably, given the current estimates for the rates of gene duplication and loss[32], the mathematical models fitted to explain power law behavior predict unrealistically long times for the generation of the largest families[30]. The time estimates become more realistic when birth and death rates among family members[30] are not independent. The adaptive radiation model implies interdependence of birth and death rates, since periods of rapid fitness landscape exploration upon the opening of new biochemical niches will involve selection for large gene amplifications in short periods of time, followed by the non-functionalization and eventual loss of the less efficient versions after a period of divergence and competition between paralogous gene copies. If this punctuated model of gene family evolution is correct, the fixation probabilities of gene duplication and loss during rapid adaptive periods will be very different from the long-term rates estimated from genome comparisons or from the silent divergence between fixed duplicate pairs[32]. This can explain the difficulty of mathematical models based on such rate estimations in accounting for the generation of the largest gene expansions in a reasonable amount of time. Just as in the case of adaptive mutagenesis, transitory bursts of gene amplification can solve the discrepancy between observed and expected rates of adaptive evolution.

2) Early Selection on Paralogous Gene Copies

In contrast to models where the early stages after gene duplication occur under neutrality, the adaptive radiation model postulates an initial period of selection for gene amplification, followed by positive selection on the paralogous gene copies for the acquisition of new function. A previous model[33] had postulated that, for bifunctional genes, positive selection immediately following gene duplication would result in each daughter gene specializing in one of the parental functions. However, the adaptive radiation model does not require multiple functions to coexist before duplication, but is rather a mechanism allowing the evolution of new functions from preexisting ones.

In bacteria, archaea and eukaryotes, extensive genome-wide analyses have supported the view that most retained paralogs never undergo substantial periods of neutral evolution[4,32,34]. Moreover, cases of positive selection shortly after duplication have been detected in analyses of sequence divergence and/or polymorphism. In mammals, adaptation via positive selection for amino acid changes has been demonstrated most convincingly for antimicrobial defensins[35], olfactory receptors[36,37], a nuclear membrane protein in humans and apes[38] and a recent duplicate of ribonuclease in a leaf-eating monkey[39]. Positive selection has also driven the divergence of recently duplicated histones and accessory gland proteins in *Drosophila*[40]. Moreover, reduced levels of nucleotide polymorphism confirm that at least one of these duplicates underwent a selective sweep within a single *Drosophila* population[40]. Fixation through positive selection is also indicated by reduced levels of nucleotide polymorphism for several recent duplications in *Arabidopsis thaliana*[41].

Most notably, evidence for early natural selection acting on gene copies that eventually became redundant and turned into pseudogenes can be found in several recent works. Some human olfactory receptor pseudogenes have patterns of nucleotide substitution indicative of purifying selection before recent non-functionalization[36,42]. Furthermore, population-level analyses show that allele frequency distributions at some of these currently pseudogenic loci are biased towards rare variants[36], suggesting the occurrence of recent selective sweeps. In addition, genome-wide analyses of gene duplication and pseudogene formation in worm, yeast and human[28], as well as in 64 bacterial and archaeal species[43], reveal that the number of duplicated pseudogenes is not proportional to the size of the gene family, also suggesting the involvement of selection in the fixation of the gene copies that eventually become pseudogenes.

3) Generation of numerous pseudogenes and eventual pseudogene loss during the establishment of new gene functions

The third general prediction of the adaptive radiation model is that "many are called but few are chosen". In this model, pseudogene formation, as well as gene and pseudogene loss, are expected during and after the occupation of a new biochemical niche, as inferior variants are outcompeted by better-adapted paralogs. And, indeed, genome-wide analyses in bacteria, archaea and eukaryotes uncover patterns that indicate an excess of pseudogene formation accompanying the successful establishment of a new gene function or group of related functions. Most remarkably, in contrast to processed pseudogenes which originate through reverse transcription of mRNA, pseudogenes generated after gene duplications tend to be associated with environmental response families[28,43], which are subject to more variable selective pressures, and are more prevalent in gene families

that have undergone organismic-specific expansions, indicative of recent selection[28]. In prokaryotes, gene families related to DNA transposition and genes acquired by horizontal transfer have higher pseudogene/gene ratios[43]. As mentioned earlier, horizontally transferred genes also undergo more duplication events[19], presumably as they adapt to the genomic environment and to the specific functional requirements of their new host.

The process of gene and pseudogene loss could simply result from genetic drift or be positively selected. Loss of gene copies that have retained the original function would likely be selected for in order to restore the dosage of the original protein before gene amplification. Similarly, loss of gene copies encoding variants of intermediate functionality and loss of pseudogenes would avoid superfluous genome size increase, expenditures in transcription and translation, and functional interference by expressed proteins of suboptimal activity. The strength of selection against these factors, along with the rate of point substitution and deletion, will determine the permanence time of superfluous gene copies in different genomes. The power law slopes of gene family size distributions are dependent on genome size, such that in large genomes the overrepresentation of the largest families increases[27]. In bacteria, it has also been noted that large genomes contain a higher proportion of genes belonging to lineage-specific gene families[22]. These correlations probably reflect the longer permanence time of large, temporary gene expansions in genomes with lower deletion rates and/or weaker constraints on genome size increase. Most dramatically, pressure to delete redundant genes and pseudogenes seems to be notably weaker in eukaryotes than in prokaryotes. In bacteria, genes that are not under sufficient selection and pseudogenes are rapidly deleted, and therefore pseudogenes are usually rare[44]. In contrast, pseudogenes are

common in most eukaryotic species. An exception is *Drosophila*, but this species contains high numbers of small protein motifs and pseudomotifs in intergenic regions[28], indicating rapid decay of pseudogenes rather than a low formation rate.

The olfactory receptor genes: a case for the adaptive radiation model?

One of the most studied cases of gene family evolution, that of the olfactory receptor genes (OR) in humans and other mammals, serves as a case example in which most of the previously discussed patterns in support of this model can be seen. The OR genes represent one of the largest mammalian gene families, with ~1000 genes in humans, arranged in clusters of up to 100 genes dispersed in more than 100 genomic locations[36,45]. Given that a high fraction (>60%) of the human OR copies has degenerated to pseudogenes, and that the human OR functional repertoire is smaller than that of other mammals, it has been assumed that the evolution of the human OR family reflects relaxed selective constraints due to our reduced reliance on the sense of smell. However, many of the characteristics of this family can be interpreted as hallmarks of an ongoing adaptation to novel odorant molecules that constantly appear in our rapidly changing lifestyle, and the evolution of this family can be seen as an example of the adaptive radiation model presented here. In particular:

1) Some OR subfamilies with functional members have duplicated so recently that their copy number, location and functionality of different paralogs are polymorphic within humans[24,25]. Thus, some human functional OR genes have appeared in the very recent past.

2) Some of the OR genes, possibly including some recent pseudogenes[36], have patterns of nucleotide substitution[36,37] and/or population

polymorphism[37,46,47] indicative of positive natural selection. This testifies to recent successful and unsuccessful attempts to produce novel functions in OR genes.

3) The fraction of OR pseudogenes in humans is high when compared to other primates and rodents[36,42]. In light of the adaptive radiation model, high pseudogenization is expected during periods of proliferation of new functions in response to environmental change.

Taken together, these evolutionary patterns strongly suggest an ongoing radiation of OR genes, rather than a simple degeneration of olfactory capabilities in the human lineage.

Integrating genic and organismal selection

The adaptive radiation model provides an opportunity for integrating different levels of selection, since the differential capacity of paralogs to survive and reproduce within the genome will influence, through the sieve of organismal fitness, which gene copies are ultimately retained. For instance, the genomic location of a sequence will influence its likelihood of undergoing point mutations, deletions or duplications, and therefore genomic rearrangements will alter the evolutionary rates of these events in translocated paralogs. As a result, the probability of incurring adaptive mutations will be altered in individuals harboring a genomic rearrangement. Eventually, selection at both the genic and organismal levels will probably have contributed to determine which gene copies were fixed in the population.

Parallels to the macroevolution of organismal lineages

Finally, this approach to the origin and diversification of gene functions rallies molecular and organismal change into a common ecological and evolutionary framework, driven mainly by adaptation to the physical and biological environment. In doing so, it broadens the application of concepts and methodologies developed in the study of organismal adaptation and radiation to the genic level. Mainly, it adopts the concepts of ecological opportunity, preadaptation, adaptive radiation and competition among close variants for a specific niche[48]. In particular, the adaptive radiation model here described presents obvious similarities to the macroevolutionary notions of quantum evolution[49] and punctuated equilibrium[50,51]. As in punctuated equilibrium, this model uses observed patterns of change within lineages (bursts of gene amplification accompanied by pseudogene formation) to deduce a hypothesis about evolutionary process (rapid change associated with lineage/gene splitting, lineage selection, and extinctions of intermediate variants due to niche competition). In terms of methodology, the vast literature developed by paleontologists and evolutionary biologists to study the rates and patterns of diversification of organismal lineages[49,51-57] and the origins of evolutionary trends[58,59] should be applicable to investigate the tempo and mode of gene family evolution, as well as the correlations of gene level events with organismal and environmental change.

Keywords: evolution of new genes, gene families, adaptive radiation, gene duplication, gene amplification, paralogs

ACKNOWLEDGEMENTS

Early discussions on the parallels between gene and organismal evolution with Mónica Medina prompted the development of this model. This manuscript was improved thanks to critical comments by Howard Ochman, Ulfar Bergthorsson, Rick Baker, Jenna Morgan, Gabriel Moreno-Hagelsieb and Elbert Branscomb. The figure was designed with input from Martha Posada-Buitrago, Yvonne Vallès, Paramvir Dehal and Eric Alm. The author thanks Fyodor Kondrashov and two anonymous referees for constructive review of the manuscript. The author also wishes to thank the U.S. Department of Energy, Office of Biological and Environmental Research, and the Joint Genome Institute, Department of Evolutionary Genomics, for support during the writing of this paper.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.

COMPETING INTERESTS STATEMENT

The author declares that she has no competing financial interests.

References

1. Ohno, S. *Evolution by gene duplication*, (Springer-Verlag, Heidelberg, Germany, 1970).
2. Force, A. et al. Preservation of duplicate genes by complementary degenerative mutations. *Genetics* **151**, 1531-1545 (1999).
3. Hendrickson, H., Slechta, E.S., Bergthorsson, U., Andersson, D.I. & Roth, J.R. Amplification–mutagenesis: Evidence that ‘‘directed’’ adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proc. Natl. Acad. Sci.* **99**, 2164–2169 (2002).
4. Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I. & Koonin, E.V. Selection in the evolution of gene duplications. *Genome Biol* **3**(2002).
5. Hooper, S.D. & Berg, O.G. On the Nature of Gene Innovation: Duplication Patterns in Microbial Genomes. *Mol. Biol. Evol.* **20**, 945–954 (2003).
6. Roth, J.R. & Andersson, D.I. Adaptive mutation: how growth under selection stimulates Lac(+) reversion by increasing target copy number. *J Bacteriol* **186**, 4855-60 (2004).
7. Cairns, J. & Foster, P.L. Adaptive radiation of a frameshift mutation in *Escherichia coli*. *Genetics* **128**, 695-701 (1991).
8. Romero, D. & Palacios, R. Gene amplification and genomic plasticity in prokaryotes. *Annu Rev Genet* **31**, 91-111 (1997).
9. Riehle, M.M., Bennett, A.F. & Long, A.D. Genetic architecture of thermal adaptation in Escherichia coli. *Proc Natl Acad Sci U S A* **98**, 525-30 (2001).
10. Zhong, S., Khodursky, A., Dykhuizen, D.E. & Dean, A.M. Evolutionary genomics of ecological specialization. *Proc Natl Acad Sci U S A* **101**, 11719-24 (2004).
11. Reams, A.B. & Neidle, E.L. Genome plasticity in Acinetobacter: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. *Mol Microbiol* **47**, 1291-304 (2003).
12. Dunham, M.J. et al. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **99**, 16144-9 (2002).
13. Guillemaud, T. et al. Quantitative variation and selection of esterase gene amplification in *Culex pipiens*. *Heredity* **83**, 87-99 (1999).
14. Kim, S.J. & Lee, G.M. Cytogenetic analysis of chimeric antibody-producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnol Bioeng* **64**, 741-9 (1999).
15. Roth, J.R. & Andersson, D.I. Amplification-mutagenesis--how growth under selection contributes to the origin of genetic diversity and explains the phenomenon of adaptive mutation. *Res Microbiol* **155**, 342-51 (2004).
16. Hall, B.G. The EBG system of E. coli: origin and evolution of a novel beta-galactosidase for the metabolism of lactose. *Genetica* **118**, 143-56 (2003).
17. Aharoni, A. et al. The 'evolvability' of promiscuous protein functions. *Nat Genet* **37**, 73-6 (2005).

18. Gevers, D., Vandepoele, K., Simillion, C. & Van de Peer, Y. Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends in Microbiology* **12**(2004).
19. Hooper, S.D. & Berg, O.G. Duplication is more common among laterally transferred genes than among indigenous genes. *Genome Biology* **4:R48**(2003).
20. Dujon, B. & et al. Genome evolution in yeasts. *Nature* **430**, 35-44 (2004).
21. Lespinet, O., Wolf, Y.I., Koonin, E.V. & Aravind, L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res* **12**, 1048-59 (2002).
22. Jordan, I.K., Makarova, K.S., Spouge, J.L., Wolf, Y.I. & Koonin, E.V. Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Res* **11**, 555-65 (2001).
23. Fortna, A. & et al. Lineage-Specific Gene Duplication and Loss in Human and Great Ape Evolution. *PLOS Biology* **2**, 0937-0954 (2004).
24. Trask, B.J. & et al. Large multi-chromosomal duplications encompass many members of the olfactory receptor gene family in the human genome. *Hum. Mol. Genet.* **7**, 2007-2020 (1998).
25. Mefford, H.C. & et al. Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes. *Hum. Mol. Genet.* **10**, 2363-2372 (2001).
26. Sebat, J. et al. Large-scale copy number polymorphism in the human genome. *Science* **305**, 525-8 (2004).
27. Huynen, M.A. & van Nimwegen, E. The Frequency Distribution of Gene Family Sizes in Complete Genomes. *Mol. Biol. Evol.* **15**, 583-589 (1998).
28. Harrison, P.M., Gerstein, M. & J. Mol. Biol. (2002) 318. Studying Genomes Through the Aeons: Protein Families, Pseudogenes and Proteome Evolution. *J. Mol. Biol.* **318**, 1155–1174 (2002).
29. Qian, J., Luscombe, N.M. & Gerstein, M. Protein Family and Fold Occurrence in Genomes: Power-law Behaviour and Evolutionary Model. *J. Mol. Biol.* **313**, 673-681 (2001).
30. Karev, G.P., Wolf, Y.I. & Koonin, E.V. Simple stochastic birth and death models of genome evolution: was there enough time for us to evolve? *Bioinformatics* **19**, 1889–1900 (2003).
31. Karev, G.P., Wolf, Y.I., Rzhetsky, A.Y., Berezovskaya, F.S. & Koonin, E.V. Birth and death of protein domains: A simple model of evolution explains power law behavior. *BMC Evolutionary Biology* **2**(2002).
32. Lynch, M. & Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
33. Hughes, A.L. The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci* **256**, 119-24 (1994).
34. Wagner, A. Selection and gene duplication: a view from the genome. *Genome Biology* **3**, 1012.1–1012.3 (2002).
35. Hughes, A.L. Evolutionary diversification of the mammalian defensins. *Cell Mol Life Sci* **56**, 94-103 (1999).

36. Gilad, Y., Bustamante, C.D., Lancet, D. & Paabo, S. Natural Selection on the Olfactory Receptor Gene Family in Humans and Chimpanzees. *Am. J. Hum. Genet.* **73**, 489–501 (2003).
37. Gilad, Y. et al. Dichotomy of single-nucleotide polymorphism haplotypes in olfactory receptor genes and pseudogenes. *Nature Genetics* **26**, 221-224 (2000).
38. Johnson, M.E. et al. Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**, 514-9 (2001).
39. Zhang, J., Rosenberg, H.F. & Nei, M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**, 3708–3713 (1998).
40. Holloway, A.K. & Begun, D.J. Molecular evolution and population genetics of duplicated accessory gland protein genes in Drosophila. *Mol Biol Evol* **21**, 1625-8 (2004).
41. Moore, R.C. & Purugganan, M.D. The early stages of duplicate gene evolution. *Proc Natl Acad Sci U S A* **100**, 15682–15687 (2003).
42. Newman, T. & Trask, B.J. Complex Evolution of 7E Olfactory Receptor Genes in Segmental Duplications. *Genome Research* **13**, 781–793 (2003).
43. Liu, Y., Harrison, P.M., Kunin, V. & Gerstein, M. Comprehensive widespread gene decay and failure of putative horizontally transferred genes. *Genome Biology* **5**(2004).
44. Mira, A., Ochman, H. & Moran, N.A. Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17**, 589-96 (2001).
45. Young, J.M. & Trask, B.J. The sense of smell: genomics of vertebrate odorant receptors. *Human Molecular Genetics* **11**, 1153–1160 (2002).
46. Gilad, Y. & Lancet, D. Population Differences in the Human Functional Olfactory Repertoire. *Mol. Biol. Evol.* **20**, 307–314 (2003).
47. Clark, A.G. Inferring Nonneutral Evolution from Human-Chimp-Mouse Orthologous Gene Trios. *Science* **302**, 1960-1963 (2003).
48. Schluter, D. *The Ecology of Adaptive Radiation*, (Oxford University Press, Oxford, UK, 2000).
49. Simpson, G. *Tempo and Mode in Evolution*, (Columbia University Press, New York, 1944).
50. Eldredge, N. & Gould, S.J. Punctuated equilibria: an alternative to phyletic gradualism. in *Models in Paleobiology* (ed. Schopf, T.) pp. 82-115 (Freeman, Cooper & Co., San Francisco, CA, 1972).
51. Stanley, S. *Macroevolution: Pattern and Process*, (W. H. Freeman, San Francisco, CA, 1979).
52. Simpson, G. *The Major Features of Evolution*, (Columbia University Press, New York, 1953).
53. Lande, R. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314-334 (1976).
54. Gingerich, P.D. Rates of Evolution: Effects of time and temporal scaling. *Science* **222**, 159-161 (1983).
55. Lynch, M. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Am. Nat.* **136**, 727-741 (1990).

56. Turelli, M., Gillespie, J.H. & Lande, R. Rate tests for selection on quantitative characters during macroevolution and microevolution. *Evolution* **42**, 1085-1089 (1988).
57. Chan, K.M. & Moore, B.R. Whole-tree methods for detecting differential diversification rates. *Syst Biol* **51**, 855-65 (2002).
58. Wagner, P.J. Contrasting the underlying patterns of active trends in morphologic evolution. *Evolution* **50**, 990-1007 (1996).
59. Jablonski, D. Body size and macroevolution. in *Evolutionary Paleobiology* (eds. Jablonski, D., Erwin, D.H. & Lipps, J.H.) 256-289 (University of Chicago Press, Chicago, IL., 1996).

Figure legends

Figure 1. Hypothetical course of events illustrating the *adaptive radiation model* for the evolution of new gene functions. The fittest genotype in the population is shown for every step. 1) The original gene locus encodes a receptor for an environmental chemical. 2) A *new chemical* appears in the environment; the receptor encoded by the original gene is *preadapted* to bind it with *very low affinity*. 3) An *amplification* of the original gene is *selectively advantageous* because it allows binding of the new chemical at biologically significant levels (assuming the benefits of binding the new chemical offset the potential deleterious consequences of binding the original one in excess). 4) In subsequent generations, different gene copies across the population will *compete* and undergo different fates; here, the fittest genotype has acquired *adaptive mutations* in two paralogs, that now encode receptors of *intermediate binding affinity* for the new chemical, as well as deleterious mutations that have turned two other paralogs into *pseudogenes*. 5) In a future generation, the fittest genotype has lost the two old pseudogenes and has aquired a new one, while one of the paralogs has undergone new adaptive mutations that confer *high binding affinity* for the new chemical. 6) Ultimately, the optimal genotype contains the original gene locus (or an identical copy), and the paralog adapted to recognize the new environmental chemical with high affinity, while pseudogenes and superfluous gene copies encoding receptors of lower affinity have disappeared. *Gene and pseudogene loss* could occur by *genetic drift* or by *positive selection*, to restore protein dosage for the original receptor and to avoid superfluous genome size expansion, expenditures in transcription and translation, and functional interference by expressed proteins of suboptimal activity.

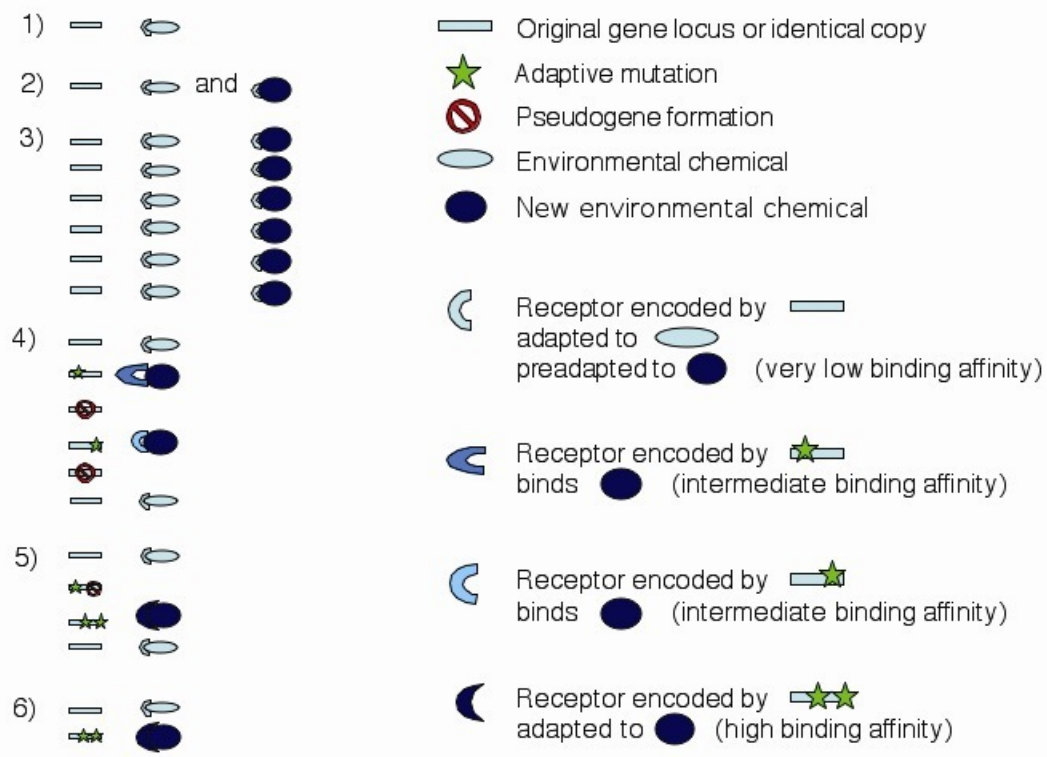
1)
2)
and
3)
4)
5)
6)
Original gene locus or identical copy
Adaptive mutation
Pseudogene formation
Environmental chemical
New environmental chemical
Receptor encoded by
adapted to
preadapted to
(very low binding affinity)
Receptor encoded by
binds
(intermediate binding affinity)
Receptor encoded by
binds
(intermediate binding affinity)
Receptor encoded by
adapted to
(high binding affinity)

Figure 1