Review TRENDS in Genetics Vol.20 No.10 October 2004

Subfunction partitioning, the teleost radiation and the annotation of the human genome

John Postlethwait, Angel Amores, William Cresko, Amy Singer and Yi-Lin Yan

Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA

Half of all vertebrate species are teleost fish. What accounts for this explosion of biodiversity? Recent evidence and advances in evolutionary theory suggest that genomic features could have played a significant role in the teleost radiation. This review examines evidence for an ancient whole-genome duplication (tetraploidization) event that probably occurred just before the teleost radiation. The partitioning of ancestral subfunctions between gene copies arising from this duplication could have contributed to the genetic isolation of populations, to lineage-specific diversification of developmental programs, and ultimately to phenotypic variation among teleost fish. Beyond its importance for understanding mechanisms that generate biodiversity, the partitioning of subfunctions between teleost co-orthologs of human genes can facilitate the identification of tissue-specific conserved noncoding regions and can simplify the analysis of ancestral gene functions obscured by pleiotropy or haploinsufficiency. Applying these principles on a genomic scale can accelerate the functional annotation of the human genome and understanding of the roles of human genes in health and disease.

From a list of all vertebrate species, pick one at random and if truly random it is just as likely that you picked a teleost fish as some other beast [1] (Figure 1). Although confined to aquatic habitats, teleosts are morphologically diverse, ranging from the smallest vertebrate, the goby *Trimmatom nanus* just 8 mm long, to the giant ocean sunfish *Mola mola*, a disk 3 m in diameter. Several factors probably contributed to this explosive burst of vertebrate biodiversity, but new genomic investigations and recent developments in evolutionary theory support the hypothesis originally advanced by Susumu Ohno [2,3] that a genome duplication event (tetraploidization) facilitated the teleost radiation. Importantly, the study of this fish-specific genome duplication is paying unexpected dividends by accelerating the investigation of conserved gene functions in human development, physiology, health and disease.

A genome duplication in the lineage of ray-finned fish

Pioneering studies revealed that zebrafish often has two ORTHOLOGS (see Glossary) of human genes [4,5] (for recent lists see Refs [6,7]). Genetic mapping studies showed that zebrafish CO-ORTHOLOGS (special types of PARALOGS) of human genes generally occupy duplicated segments on different zebrafish chromosomes [8–17]. Did most zebrafish duplicated chromosome segments arise by segmental duplication as in the human genome [18], or by genome duplication [19]? Consider, for example, human chromosome 17 (Hsa17): among 74 zebrafish orthologs of Hsa17 genes mapped in zebrafish [12], at least 15 (20%) were present in duplicate, with one copy on zebrafish linkage group (LG) LG3 or LG5, and the other on LG12 or LG15; furthermore, most single-copy orthologs from *Homo sapiens* chromosome 17 (Hsa17) were also on these chromosomes [12]. A similar pattern holds for Hsa2q and LG9 plus LG6, Hsa9 and LG5 plus LG2 and LG21, Hsa10q and LG12 plus LG13 and other chromosomes [9–13]. The duplication of entire chromosomes or long chromosome arms is more likely to have occurred by whole genome duplication than segmental duplications. Because independent tetraploidy events have happened recently and repeatedly in ray-finned fish such as salmonids, goldfish and carp [20–22], the most parsimonious explanation is a whole-genome duplication event in the ancestry of zebrafish [8,9,23,24]. We do not know if this occurred as an ALLOTETRAPLOID event, as in many plants [25] and probably carp [22], or an AUTOTETRAPLOID event. These two hypotheses can be distinguished by phylogenetic or karyotypic analyses that discriminate between one or two genomes as the founders of the tetraploid [22,25]. Unfortunately, the antiquity of the event under discussion (see below) will probably frustrate resolution of this issue. Regardless of the precise mechanism of genome duplication, the outcome is the same: for at least 20% of human genes, zebrafish has two co-orthologs.

When did genome duplication occur in the lineage of ray-finned fish?

The same early work that provided a mechanism for the expansion of zebrafish gene families also provided evidence as to when the duplication event took place [8,9]. Amores *et al.* [8] showed that among the four pufferfish (*Fugu*: *Takifugu rubripes*) *Hox* clusters known at the time [26], one was orthologous to the zebrafish *hoxaa* cluster,

Corresponding author: John Postlethwait (jpostle@uoneuro.uoregon.edu).
Available online 19 August 2004

 0168-9525/$ - see front matter doi:10.1016/j.tig.2004.08.001

Glossary

Allotetraploid: a hybrid individual with two haploid genomes from one species and two haploid genomes from a different species.
Autotetraploid: an individual with four haploid genomes all from the same species.
Co-orthologs: a set of at least three genes in two species that are derived from a single gene in the last common ancestor of the two species, followed by gene duplication in the lineage of at least one of the species after the two species diverged. Co-orthology represents a special case of paralogy, but is more information-rich because it indicates the position of gene duplication with respect to the splitting of lineages.
Haploinsufficiency: a situation in which an individual that is heterozygous for a deletion mutation or null activity mutation shows a phenotype because a single copy of the normal gene is insufficient for normal function, for example, by not providing sufficient protein to ensure a normal phenotype; thus, mutations for haploinsufficient loci are dominant.
Neofunctionalization: a term introduced [49] to describe a possible mechanism for preserving both duplicates of a gene when 'one copy may acquire a mutation conferring a new function, which becomes fixed through positive Darwinian selection'. Preserved genes are no longer subject to loss because the functions of both genes have become essential. Beneficial mutations can also occur after duplicate preservation by other means or in nonduplicated genes.
Nonfunctionalization: a term coined [49] to indicate when, after gene duplication, 'one copy may incur a null mutation in the coding region, which subsequently drifts to fixation, leading to gene loss'. The other copy will generally remain intact and perform the original functions of the preduplication gene. A gene duplicate lost after nonfunctionalization can suffer additional mutations that block transcription or translation and become a pseudogene, which eventually can decay beyond recognition.
Orthologs: A pair of genes, one in each of two different species, which are descended from a single gene in the last common ancestor of the two species. Note that orthologs are defined by their history, not by their function.
Paralogs: two or more genes within a lineage that are derived from the duplication of a single ancestral gene.
Pleiotropy: a situation in which a single gene controls several distinct and sometimes otherwise seemingly unrelated phenotypic traits. For example, mutations in the human gene *SOX9* cause male-to-female sex reversal and hypoplastic scapula.
Quantitative subfunction partitioning: mutation-induced decrease in the level of function of two gene duplicates such that the sum of both co-orthologs is required to achieve the ancestral function.
Rediploidization: after a genome duplication, four homologous chromosomes align in meiosis and any can recombine with any other, which sometimes produces aneuploid gametes. In rediploidization, there arise chromosome rearrangements (e.g. inversions) that inhibit pairing and recombination of duplicate chromosomes, resulting in two pairs of homologous chromosomes that subsequently assort regularly in a mendelian fashion.
Spatial subfunction partitioning: the reciprocal degenerative mutation of tissue-specific regulatory elements such that co-orthologs are expressed in different, but perhaps overlapping, sets of tissues.
Subfunction: a specific subset of a gene's function, either regulatory or coding, that, if mutated, establishes a distinct complementation group in genetic crosses (see, e.g., Ref. [94]).
Subfunctionalization: acording to Force et al. [49], a possible mechanism for preserving both duplicates of a gene in which 'each duplicate may experience loss or reduction of expression for different subfunctions by degenerative mutations. In such a case, the combined action of both gene copies is necessary to fulfill the requirements of the ancestral locus'.
Subfunction partitioning: the segregation of ancestral gene subfunctions between two gene duplicates, whether that partitioning was responsible for duplicate preservation (subfunctionalization) or occurred after the preservation event [49]. In contrast, 'divergent resolution' has been defined as the loss of different gene duplicates in different populations [73].
Temporal subfunction partitioning: the reciprocal degenerative mutation of time-specific regulatory elements such that co-orthologs are expressed in the same tissue, but at different times under complementary regulatory elements.

one to the *hoxab* cluster and the other two to the *hoxba* and *hoxca* clusters. This implied that the event that produced duplicate copies of zebrafish *hox* clusters occurred before the divergence of zebrafish and *Fugu* lineages, and that additional *Fugu Hox* clusters were either missing or remained to be described [8,27]. Subsequent work isolated the remaining pufferfish *hox* clusters and confirmed their duplicated nature [28–30]. Additional work showed that *Fugu* had duplicated sections of Hsa20q [31,32], and comparative genomics showed that these duplications occurred before the divergence of *Fugu* and zebrafish lineages [15]. The availability of the draft sequence of the *Fugu* genome revealed many additional examples of genes duplicated before the divergence of zebrafish and pufferfish lineages [33–41]. Two recent careful global analyses of gene duplication in the *Fugu* genome show that *Fugu* has at least several hundred duplicated chromosome regions, confirming conclusively that a genome duplication occurred in the *Fugu* ancestry [42,43]. Estimates based on molecular clocks and phylogenetic analysis accounting for mutational saturation [16] suggest that the genome duplication event occurred ~350 million years ago, which is probably before the beginning of the teleost radiation [17,42–44],

Additional evidence that genome duplication preceded the teleost radiation comes from analysis of the content of whole chromosomes. Recent comparative mapping shows that medaka chromosomes have clear orthologs in zebrafish [45], as would be expected if the genome duplication occurred early in the teleost radiation (Figure 1). For example, zebrafish chromosomes LG3 and LG12 [9,12,13], and medaka chromosomes LG8 and LG19 [45] are both copies of most of human chromosome 17 (Hsa17), with additional material orthologous to portions of Hsa6, Hsa16, Hsa19 and Hsa22. In addition, zebrafish LG12 has one of the two duplicates of much of Hsa10 (the other being zebrafish LG13 and its medaka ortholog LG15). The most likely model to explain these results is that an ancient ray-finned fish had a single chromosome with material now on Hsa6, 17, 16, 19 and 22, with (or alternatively, without) the long arm of Hsa10 (Figure 1b). This chromosome duplicated, and then the Hsa10 portion translocated to (or alternatively, away from) one member of the pair. Finally, the zebrafish and medaka lineages separated, and since then these chromosomes have remained little changed except for inversions, consistent with the relative uniformity of teleost karyotypes [45].

Work with deeply diverging teleosts, such as eels and herring, and with the latest diverging non-teleosts, including bowfin, paddlefish and gar pike [46], should be undertaken to pinpoint the timing of the genome duplication event accurately. The bichir, a deeply diverging ray-finned fish, appears to have branched from other ray-finned fish before the genome duplication [47].

Evolution of gene duplicates in teleost fish

The teleost genome duplication presents previously unanticipated advantages for the analysis of gene function because of principles that govern the evolution of gene duplicates. After genome duplication, each gene copy can follow a separate evolutionary trajectory, little affected by unequal recombination or gene conversion, which homogenize tandem duplicates [48]. In the classical model [2], new gene duplicates face one of two fates: either one copy mutates to a pseudogene (called NONFUNCTIONALIZATION in Ref. [49]), or one copy preserves the original function, and the other copy mutates freely until by chance it obtains a mutation that confers a new, beneficial, positively selected function (called NEOFUNCTIONALIZATION in Ref. [49]). In a third possibility called SUBFUNCTIONALIZATION [49], the

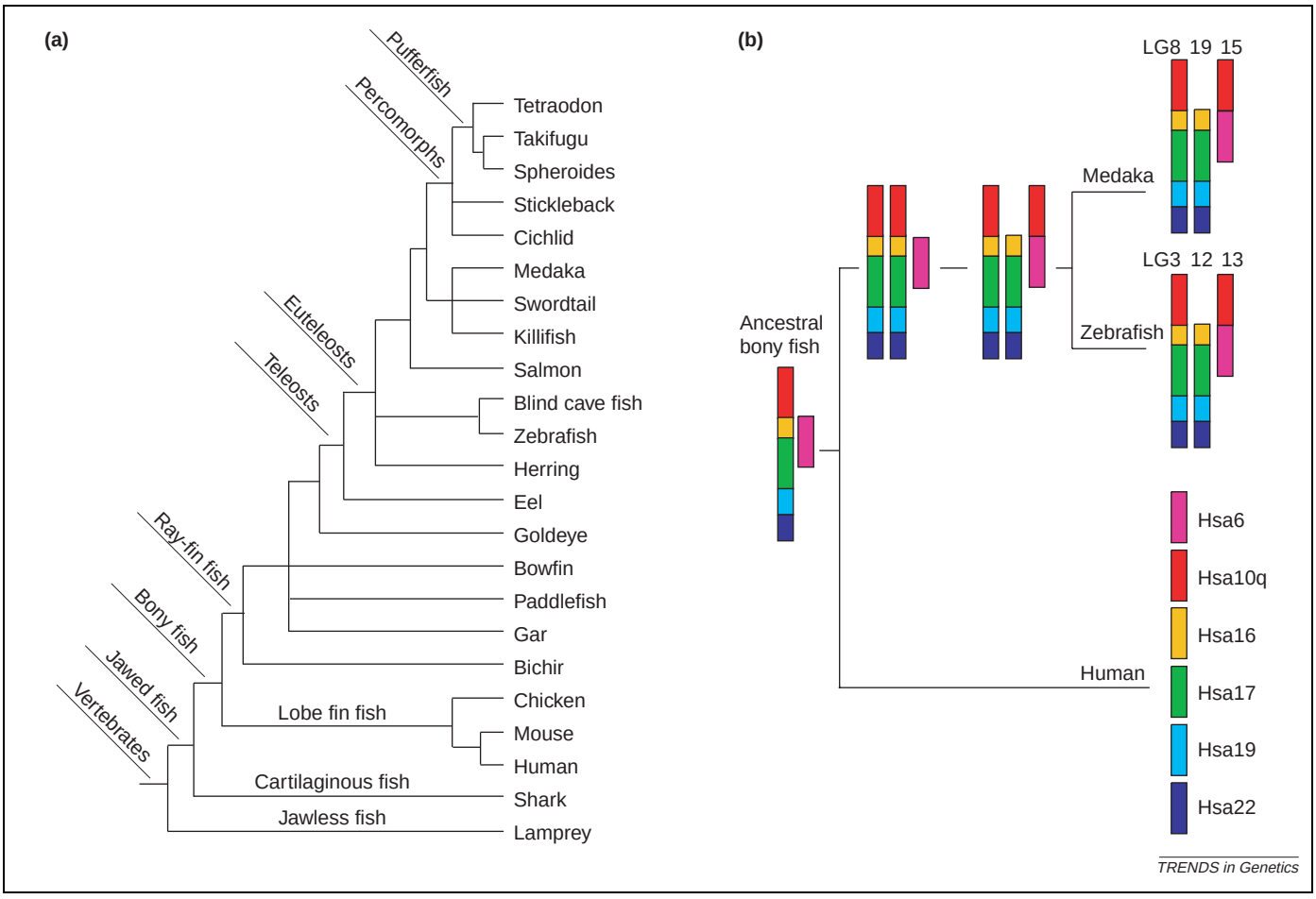

Figure 1. A phylogeny for vertebrates and some fish chromosomes. **(a)** A series of dichotomous branches characterizes a widely accepted hypothesis for the vertebrate tree [1,46]. The first fork splits the basally diverging jawless fish, including lamprey and hagfish, from the jawed fish. Jawed fish have two main branches: the cartilaginous fish, including sharks and rays; and the bony fish, which, in contrast to the earlier diverging lineages, generally have bony skeletons. Bony fish diverged again into two main lineages, the lobe-finned fish, such as humans, chickens and lungfish, and the ray-finned fish, including sturgeon and trout. As their names suggest, a general feature that distinguishes ray-finned and lobe-finned lineages is the construction of their paired appendages. Several lineages diverged basally among ray-finned fish, including bichir, sturgeon and bowfin. The crown group among ray-finned fish is a large division called the teleosts. Teleost-specific skeletal features, including the structure of the neural arch of the vertebrae and specializations of the oral skeleton, can improve teleost mobility and predatory effectiveness. Thick lines indicate a likely time of the genome duplication event. **(b)** Medaka has orthologs of duplicated zebrafish chromosomes, which is strong evidence of a genome duplication event before the divergence of medaka and zebrafish lineages. For example, the medaka LGs 8, 19 and 15 are orthologus to zebrafish LGs 3, 12 and 13. A portion of a fish chromosome orthologous to Hsa10q either translocated away from the precursor of medaka LG19 and zebrafish LG12, as shown in the figure or, alternatively, translocated from another chromosome to the precursor of medaka LG8 and zebrafish LG3 (not shown for simplicity). In either explanation, this event must have occurred after the genome duplication but before the divergence of zebrafish and medaka lineages.

complementary partitioning of ancestral regulatory and structural SUBFUNCTIONS is the mechanism that preserves gene duplicates (see Ref. [50] for related ideas). Note that the mechanism that initially preserves a pair of duplicates is independent of subsequent acquisition of novel functions and SUBFUNCTION PARTITIONING.

Nonfunctionalization is the fate of 50–90% of gene pairs duplicated in animal polyploidization events [51]. Defining the precise mechanism of duplicate preservation from current patterns of gene expression and function, however, is difficult. Evidence that one member of a gene duplicate has acquired a novel function does not necessarily mean that this positively selected function actually caused the initial preservation of the gene (neofunctionalization). It is possible that the new function evolved subsequent to the duplicate preservation event.

There are several possible examples of neofunctionalization in fish. Teleosts have two copies of the human estrogen receptor gene *ESR2*. The *esr2b* co-ortholog diverged rapidly in sequence shortly after the duplication event, followed by a much slower rate of sequence change subsequently [52]. This suggests that novel functions emerged in *esr2b* shortly after the duplication event, and that this constrained further divergence of the *esr2b* sequence, whereas the sequence of *esr2a* continuously drifted slowly [52]. Two other cases involve *mitf* [33,53] and *sox9* [54], where one gene duplicate is expressed in tissues (epiphysis and ovary, respectively) where the ortholog in an outgroup, the mouse, is not expressed [55,56]. In both cases, the fish-specific expression pattern could result either from the evolution of a novel, positively selected subfunction in one of the zebrafish duplicates, or from the loss of an ancestral subfunction in the mouse lineage. Genomic resources and embryological material for the most recent nonduplicated ray-finned fish would help to resolve such issues. Note that both *mitf* and *sox9* gene duplicates show subfunction partitioning in addition to what might be the origin of novel functions.

Evidence for subfunction partitioning

Although it can be difficult to demonstrate that any particular pair of gene duplicates was retained by subfunctionalization, subfunction partitioning is common among duplicated genes arising from the preteleost genome duplication event. Subfunction partitioning can involve regulatory and/or structural subfunctions and has important evolutionary consequences. Assume, for instance, that an ancestral gene is expressed in liver and brain from independently mutable regulatory elements, and that specific domains of the protein interact with tissue-specific coactivators (Figure 2). After duplication, subfunctions can reciprocally degenerate in duplicates, leading to co-orthologs with tissue-specific essential functions, a case of SPATIAL SUBFUNCTION PARTITIONING.

As an example, consider *eng1a* and *eng1b*, zebrafish co-orthologs of mammalian *EN1*. The fish co-orthologs are expressed in the limb bud and hindbrain, respectively [4,49]. By contrast, mouse *En1* is expressed in both tissues [57]. Because expression of the fish duplicates sum to the mouse pattern, *En1* in the last common ancestor was probably expressed in both organs, and degenerative mutations destroyed complementary regulatory subfunctions in the fish genes (Figure 2). Subfunction partitioning can be qualitative (like *En1*), quantitative or temporal [49,50].

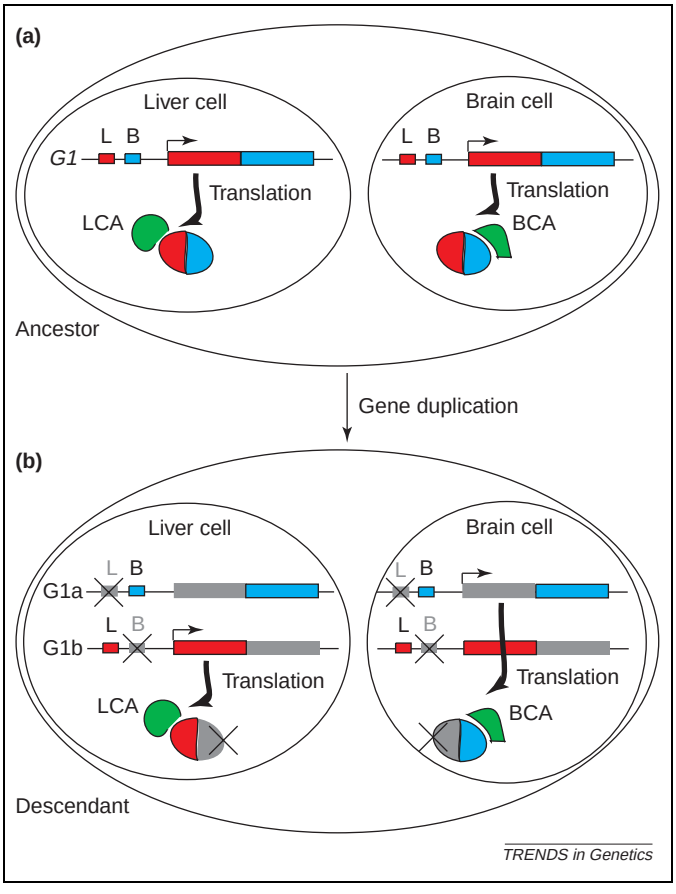

Figure 2. Partitioning of regulatory and structural subfunctions. **(a)** Ancestral gene *G1* is expressed in liver and brain from independent regulatory elements L and B. The long bar represents the coding region, with red and blue portions encoding domains of the G1 protein that interact with tissue-specific coactivators (green) in the liver (LCA) or brain (BCA). A haploid genome is shown for simplicity. **(b)** After duplication and degeneration, subfunctions (red and blue) of the gene and/or protein can reciprocally degenerate in duplicates *G1a* and *G1b*, the co-orthologs of gene *G1*.

In QUANTITATIVE SUBFUNCTION PARTITIONING, degenerative mutations can decrease the amount of product so that the sum of both genes is required to reach the ancestral function. In TEMPORAL SUBFUNCTION PARTITIONING, both co-orthologs can be expressed in the same tissue, but at different times under different regulatory elements. Protein subfunction partitioning involves complementary loss of essential protein domains such that both protein products are required for survival (Figure 2). In each case, the sum of duplicate gene functions equals the function of the ancestral preduplication gene, and any differences should be caused by the acquisition of novel gene subfunctions or the loss of ancestral functions in the outgroup.

Examples of subfunction partitioning abound in zebrafish, medaka, *Xiphophorus* and stickleback [58–70]. Despite a large list of examples, in only a few cases has the complementary sharing of mammalian gene functions by zebrafish duplicates been explained on a molecular genetic level. In one example, the mouse *Mitf* gene is transcribed from tissue-specific distal and proximal promoters [55,71]. Both zebrafish and *Xiphophorus* have two *MITF* co-orthologs, one using the distal and the other the proximal promoter [33,53]. This gives two protein isoforms as in mouse, but from two duplicated genes rather than one gene. In another example, zebrafish has two co-orthologs of mammalian *HOXB1* [5,8], one with early expression and the other with later, stable expression, a case in which temporal subfunctions have been partitioned [67]. The regulatory elements defined in mouse *Hoxb1* are conserved in *Fugu* and zebrafish [72], but one zebrafish duplicate has lost the early element, and the other duplicate has lost the late element [67]. These might currently be the only examples in which subfunction partitioning is understood at a molecular genetic level, and there is a strong need for concerted action to identify the molecular basis for subfunction partitioning in many cases to be able to generalize about the mechanisms of this phenomenon.

Nonfunctionalization, subfunction partitioning and lineage divergence

What has been the role of duplicate gene evolution in the teleost radiation? Different genes could have experienced nonfunctionalization in different teleost lineages (called divergent resolution in Ref. [73]). For example, the zebrafish lineage maintained two copies of the *hoxc* complex, but the pufferfish and medaka lineage retained just a single copy of the *hoxc*complex, whereas the reverse is true for the *hoxd* complex [8,29,30,45]. An open question is to what extent such lineage-specific nonfunctionalization is responsible for lineage divergence and the evolution of lineage-specific morphologies and behaviors.

Divergent resolution and lineage-specific subfunction partitioning can foster incompatibility among populations within a species, and thus might facilitate evolutionary radiation [74]. Imagine that regulatory subfunctions drive expression of a particular gene in the liver and brain (Figure 3). After a genome duplication event, the co-orthologous chromosomes bearing this gene become rediploidized (Figure 3a). REDIPLOIDIZATION blocks

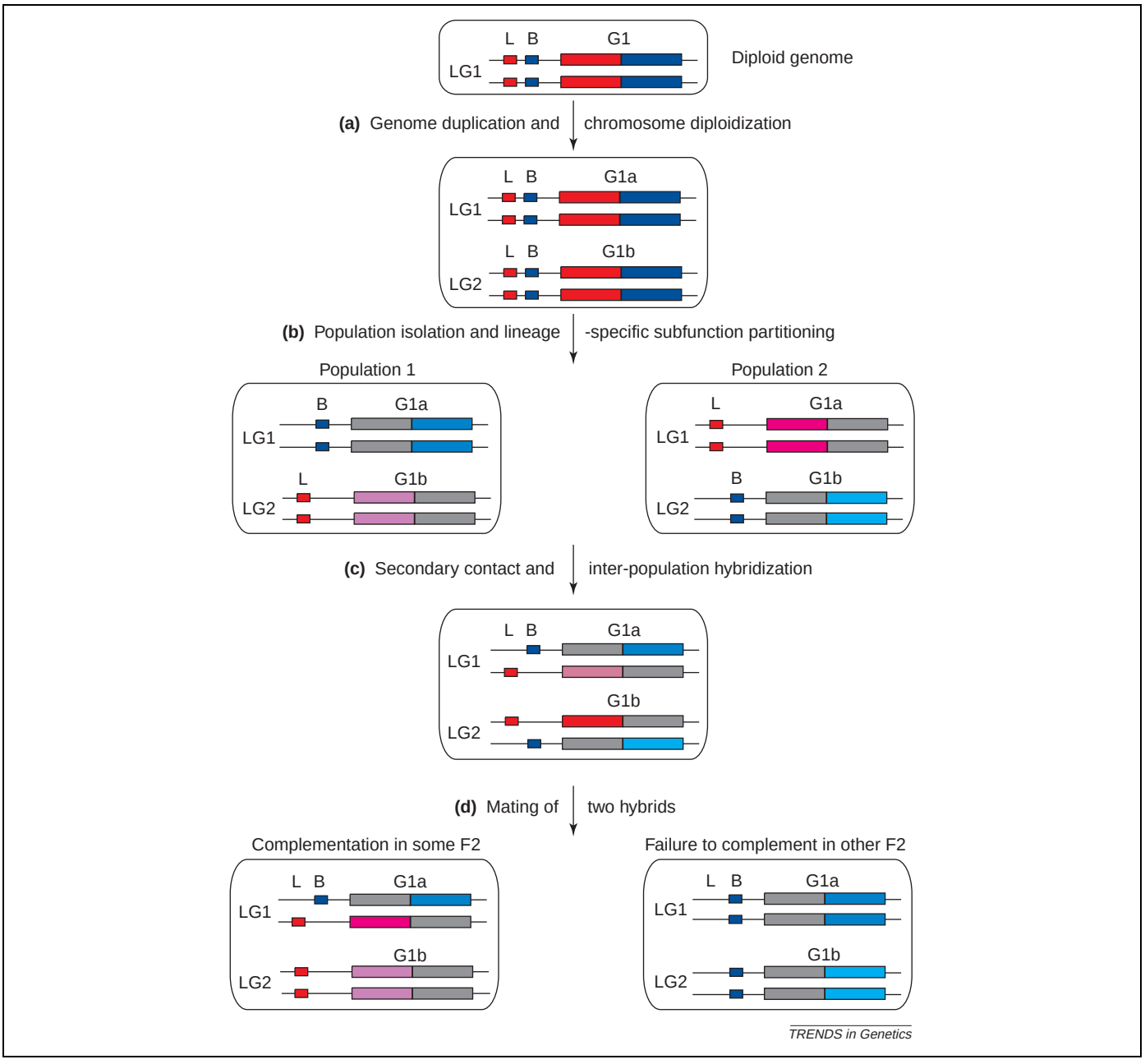

Figure 3. A mechanism for the divergence of populations leading to speciation through subfunction partitioning. Part **(a)** shows a single gene on a single pair of homologous chromosomes in a diploid ancestral species. After genome duplication and chromosome rediploidization, subsequent neutral processes of mutation and genetic drift can lead to subfunction partitioning and differentiation of the ancestral linkage group into LG1 and LG2. **(b)** If isolated populations form in this rediploidized species, then purely degenerative mutations followed by genetic drift can lead to the partitioning of subfunctions to different co-orthologs and their subsequent fixation in the two populations. **(c)** Hybrids between the two species should in general develop normally but when two hybrids mate **(d)**, a portion of their offspring will be homozygous for alleles lacking one essential subfunction (1/16th of the F2) or the other (another 1/16th), thus reducing the fitness of hybridizing individuals. If this happens for many or all of the chromosomes in a species derived from genome duplication, then hybrids could be nearly infertile. The reduction in hybrid fitness between the two populations could foster lineage divergence, reduction in gene flow and speciation through reinforcing selection on mate recognition systems. A mechanism such as this could have contributed to the teleost radiation.

recombination between gene copies on the two duplicated chromosomes, thereafter, duplicated genes can begin to evolve independently. Now imagine that two descendant populations temporarily become geographically isolated, and that subfunction partitioning occurs differently in gene copies in the two homeologous chromosomes (Figure 3b; see also Ref. [73]). Note that this lineage-specific subfunction partitioning results in true orthologs in two populations having different functions. If individuals from the two differentiated populations now come into contact and mate, the resulting hybrids should develop correctly because each subfunction is performed by one of the genes from each species (Figure 3c). The hybrids will give four types of gametes as in a mendelian dihybrid cross, so most of the F2 zygotes will have both the liver and brain subfunctions. Note, however, that 1/16th of the F2 will lack the liver subfunction, and an additional 1/16th will lack the brain subfunction (Figure 3d). This 1/8th of the F2 progeny will die if the subfunctions are essential, leading to diminished hybrid fitness. Now imagine that this happens for at least one gene on every chromosome, which would be likely in a genome

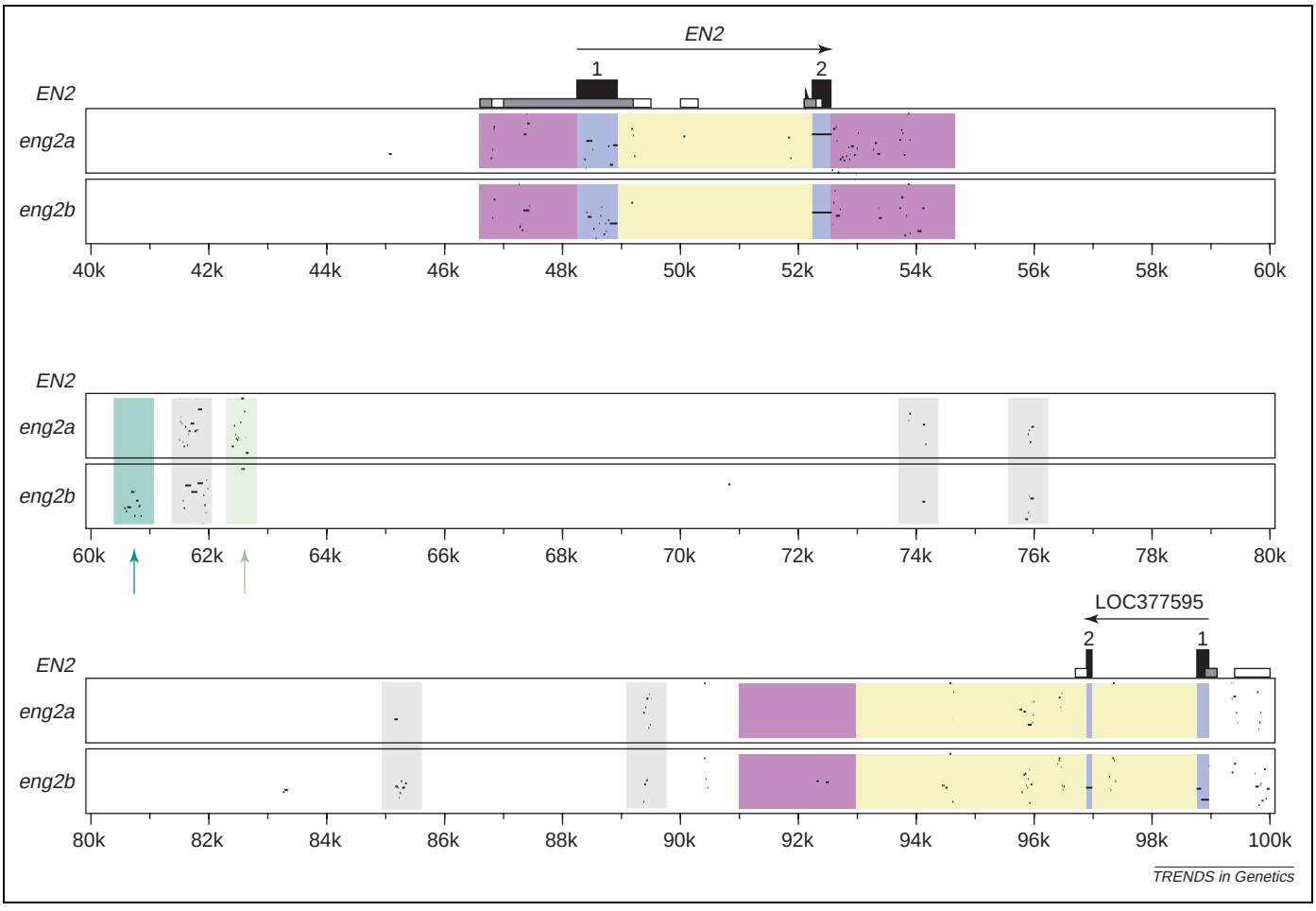

Figure 4. Subfunction partitioning and conserved noncoding regions. A percent identity plot (PIP) [80] for the *EN2* co-orthologs in human and zebrafish shows the sequence surrounding the human *EN2* gene across the top, and *LOC377595* is the adjacent locus in human and zebrafish. The horizontal band labeled *eng2a* compares a sliding window of *eng2a* sequence to the human sequence and, at the indicated position in the human gene, places a dot at a height proportional to percent identity, ranging from 50% at the bottom to 100% at the top of the band. Exons are shown in blue vetical bars, 3′ and 5′ untranslated portions of the mRNA in purple, and introns in yellow. The intergenic region between *EN2* and *LOC377595* contains several conserved noncoding (CNC) regions shared by both zebrafish duplicates shown in gray, and two that appear to have experienced subfunction partitioning, shown in dark green and light green. The dark green region is a candidate for a regulatory element essential for the anterior hindbrain expression domain and the light green region is a candidate for a regulatory element necessary for the somite expression domain, tissues that are expressing *EN2* and the indicated zebrafish co-orthologs. Targeted functional analyses of such domains could accelerate functional annotation of the human genome.

duplication event. The ensuing reduction in hybrid fitness might lead to speciation through reinforcing selection on mate recognition systems to reduce interpopulation matings [74]. By this mechanism, lineage-specific subfunction partitioning could accelerate rates of teleost speciation and lead to the diversification of characters used in mate recognition systems, which themselves might be important for adaptation to local ecological conditions [75].

Has subfunction partitioning occurred differently in various teleost lineages?

To evaluate the extent to which diverse teleost lineages have actually partitioned ancestral subfunctions differentially, we must examine the functions of both co-orthologs in more than one teleost. Gene expression patterns can serve as surrogates for regulatory subfunctions, but unfortunately, we know the detailed expression patterns of both duplicates in two different teleosts in very few cases. For *MITF*, most subfunctions appear to have partitioned before the divergence of zebrafish and *Xiphophorus* lineages [33,53]. The same is true for the aromatase gene *CYP19*; in humans this is expressed in the brain and ovary, but in many teleosts, *cyp19a* is expressed in the ovary and *cyp19b* is expressed in the brain [65,76]. For teleost duplicates of human *SOX9*, by contrast, some expression pattern evolution occurred after the divergence of lineages [68]. The finding that screens for mutations that affect embryonic development recover an overlapping but different spectrum of phenotypes in zebrafish and medaka [77] would be expected according to the hypothesis that many subfunctions were partitioned differently between the two species. Although scarce, initial data suggest that most subfunctions could have partitioned before the teleost radiation, but some appear to have partitioned afterwards. This points out that a gap in our resources is knowledge of gene expression patterns and gene abrogation phenotypes for both duplicated gene copies for many genes in phylogenetically distant teleosts.

Subfunction partitioning can help identify genetic regulatory elements

Subfunction partitioning of teleost genes provides a special opportunity to identify tissue-specific regulatory elements. Because of the antiquity of their divergence,

teleosts and tetrapod genome sequences have randomized except where function constrains sequence. Thus, conserved noncoding (CNC) sequences suggest functional genomic elements [78], and these sequences can be verified in functional tests [79]. Convenient computer programs are now available to identify CNC elements [80,81]. When applied to teleost co-orthologs that have partitioned regulatory subfunctions as in Figure 2, and coupled with known gene expression patterns, this methodology has the potential to suggest candidates for tissue-specific regulatory elements. For example, in mouse, *EN2* is expressed in the midbrain–hindbrain border, anterior hindbrain, jaw muscles and somites [82]. For the fish co-orthologs *eng2a* and *eng2b,* expression overlaps in parts of the midbrain-hindbrain border and the jaw muscles, but *eng2a* is expressed in the somites, and *eng2b* is expressed in the anterior hindbrain, as would be expected from subfunction partitioning [4]. Percent identity plots (PIP) [80] reveal noncoding sequences conserved between human and zebrafish *EN2* genes (Figure 4). CNC regions shared by both *eng2a* and *eng2b,* such as the one at 62k in Figure 4, are candidates for regulatory elements that control expression domains shared by all three *EN2* co-orthologs, such as part of the midbrain-hindbrain junction. Others, like the CNC sequences at 60.8k and 62.5k, are candidates for tissue-specific regulatory elements affected by subfunction partitioning, in this case, for expression in the anterior hindbrain for *eng2b* and the jaw for *eng2a*. Functional analyses are required to test the predictions inferred by the coupled PIP-subfunction partitioning method, but its widespread use could greatly accelerate the functional annotation of the human genome.

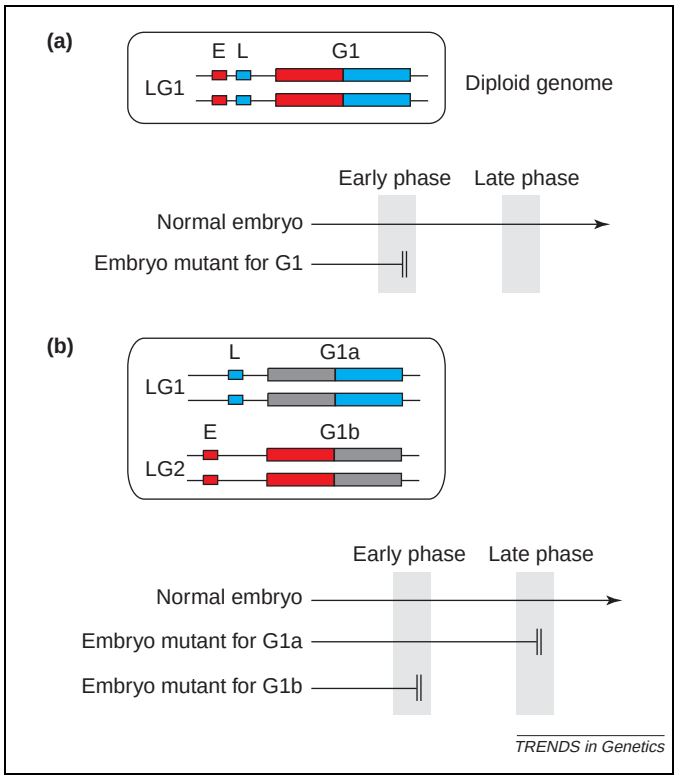

Figure 5. Subfunction partitioning facilitates the identification of conserved gene functions hidden by pleiotropy. **(a)** If an ancestral gene is required at multiple times during development, then a mutation in that gene can obstruct development at the earliest stage the gene is necessary, and the embryo might never develop to the stage that requires the later function. **(b)** After genome duplication and the partitioning of temporal subfunctions, however, animals homozygous for a mutation in a co-ortholog essential for the early subfunction will again be disrupted at the early time point, but an embryo with a mutation in the other gene copy will sail through the early critical period, and arrest at the later stage that needs the gene. Late ancestral pleiotropic subfunctions in mammalian genes will thus be hidden, but subfunction partitioning can readily expose these late subfunctions to the teleost developmental geneticist.

Subfunction partitioning in teleosts can facilitate analysis of gene function

Besides providing fundamental insights into the evolution of gene function, the origin of biodiversity, and analysis of conserved noncoding DNA, subfunction partitioning provides distinct advantages for functional genetic analyses. These include the identification of gene functions obscured in mammals by either PLEIOTROPY or HAPLOINSUFFICIENCY.

Many genes are essential at several developmental stages. Null mutations in such genes will block development at the stage of the earliest essential subfunction and this will probably obscure the later role of the gene. By contrast, the partitioning of conserved temporal subfunctions between teleost duplicates can reveal both gene subfunctions in loss-of-function experiments (Figure 5). An example of this principle involves *Nodal*. Mouse embryos homozygous for a *Nodal* null allele are blocked in early gastrulation [83], which masks identification of later functions in these animals. Zebrafish has two *Nodal* genes called *cyclops* and *squint* [58,84–86]. Single mutations and double mutant combinations in these genes show that one acts early in mesoderm induction, and the other later in neural-plate patterning. The late function is obscured by pleiotropy in the targeted mouse mutation because of developmental arrest at gastrulation, before the late function occurs. The phenotype of the zebrafish double mutant mimics that of the mouse *Nodal* knockout. Another example concerns *SOX9*. In tetrapods, loss of *Sox9* function leads to the complete loss of endochondral-derived craniofacial skeletal elements [87,88]. Zebrafish has two co-orthologs of *SOX9*, and embryos homozygous for a mutation in one of them form craniofacial cartilages in which the chondrocytes fail to stack [89]. The role for *Sox9* in chondrocyte stacking is obscured in tetrapods by pleiotropy because early *Sox9* function (which in zebrafish partitioned to *sox9b*) is required for the formation of the cells that stack. The investigation of gene abrogation phenotypes in both teleost co-orthologs of additional mammalian genes should reveal additional instances of conserved subfunctions that have not been apparent in targeted mutations in mouse because of pleiotropy.

Subfunction partitioning also facilitates analysis of mammalian genes that are haploinsufficient. Mammalian heterozygotes for null alleles in *SOX9* usually die because of hypoplastic cartilage [90]; apparently a single dose of the normal allele of *SOX9* delivers insufficient gene product to provide a normal phenotype. Because heterozygotes die as infants, it is not possible to obtain homozygous mouse embryos by the mating of two heterozygotes. In addition, because mutants heterozygous for null alleles generally have about half the usual level of normal gene activity, they

Box 1. Subfunction partitioning and drug discovery

Subfunction partitioning of teleost gene duplicates could be useful for drug discovery. Some proteins (e.g. protein G1 in Figure Ia) that are targets of therapeutic drugs (shown in purple) can interact with different cofactors in different cell types, for example, LCA, the liver coactivator, or BCA, the brain coactivator. The intended drug target might be the function of protein G1 in a single cell type, for instance, liver cells, but the drug could alter the activity of G1 everywhere it is found leading to undesirable side effects.

After the teleost gene duplication and subfunction partitioning (Figure Ib), the liver-specific coding domain could have degenerated in the brain-expressed gene duplicate (*G1a*), and reciprocally, the brain-specific domain could have degenerated in the liver-specific co-ortholog (*G1b*). In such a case, the liver-specific teleost protein would rescue the liver phenotype in a mammal mutant for the unduplicated ortholog, but it would not rescue the brain phenotype, and vice versa for the brain-specific teleost protein. This would identify tissue-specific protein domains.

The identification of tissue-specific protein domains from the subfunction partitioning of teleost gene duplicates could help guide the design of drugs targeted to a single cell type if one could design or identify a drug that specifically affects just one of the identified tissue-specific protein domains (Figure Ib right). A drug targeted to a tissue-specific protein domain could alleviate side effects that arise from inhibiting the protein in the wrong tissue. In some cases, teleost genes could be directly effective in the treatment of human disease. Salmon calcitonin is commonly used in the treatment of bone diseases such as Paget's disease, hypercalcemia and osteoporosis because it has fewer side effects than the mammalian protein [95]. Because subfunctions might have partitioned differently in different teleost lineages, it will be useful to examine several teleosts to find different tissue-specific protein domains.

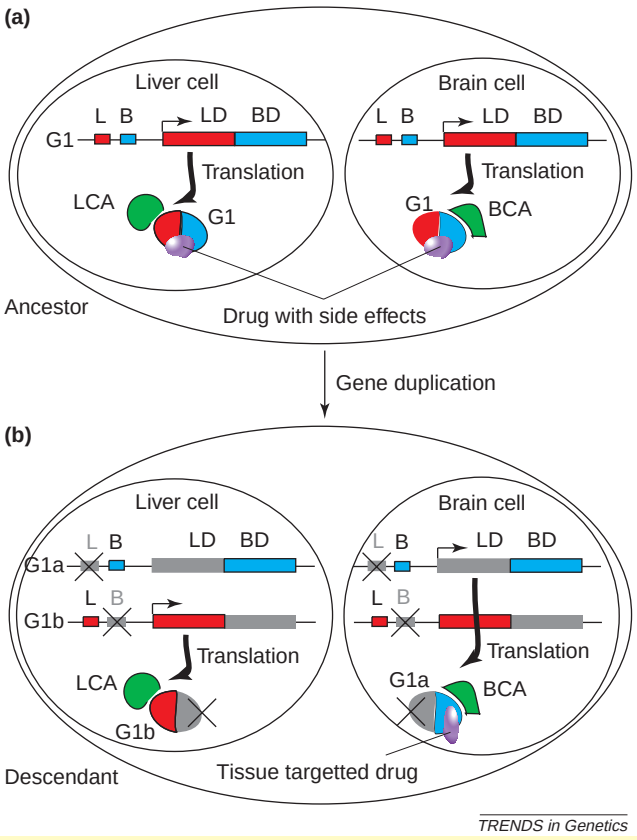

Figure I. The application of subfunction partitioning of teleost gene duplicates for drug discovery.

do not fully reveal the functions of the mutated gene. Subfunction partitioning, however, can distribute a gene's roles between co-orthologs, which might allow heterozygotes to survive. Teleosts have two copies of *sox9*, and mutations in them behave as recessive lethals rather than dominant lethals as in mouse [89]. Phenotypic analysis of embryos homozygous for *sox9a* revealed that *sox9* is essential for chondrocyte stacking and the production of cartilage. In mouse, conditional alleles or genetic mosaics must be constructed to obtain mice homozygous for a *Sox9* knockout [88,91]. In general, identification of ancestral functions of loci that are haploinsufficient in humans and mouse will probably be more efficient in teleosts because of subfunction partitioning.

Conclusions

Recent evidence converges on the conclusion that a genome duplication event preceded the grand radiation of teleosts into the most species-rich group of vertebrates. Subfunction partitioning could be a universal evolutionary pathway followed by co-orthologs resulting from this event, and perhaps by duplicated genes in general. It is still unclear how often teleost co-orthologs have partitioned subfunctions and how often they have evolved novel functions. A consequence of subfunction partitioning suggested by evolutionary theory is the isolation of populations through reduction in hybrid fitness, which can potentially lead to speciation. More work is needed in this area, but recent data suggest that lineage-specific subfunction partitioning has indeed occurred [68]. An outstanding problem is the relative amount of subfunction partitioning that took place before and after the divergence of teleost lineages.

Subfunction partitioning provides advantages for identifying the conserved roles of single-copy mammalian genes because of the nature of conserved noncoding regions, the relaxation of pleiotropy, and the loss of haploinsufficiency. Because subfunctions appear to have partitioned differently in different lineages, all these analyses would benefit by expanding analysis from zebrafish, pufferfish and medaka to other teleost models such as those shown in Figure 1.

A crucial tool for the interpretation of functional analyses of teleost duplicates is an appropriate outgroup. Mouse is well studied and useful, but a fish occupying the most recently diverging nonduplicated lineage is indispensable. Substantial effort should be directed to evaluating the genomes and embryos of bowfin, paddlefish and gar pike to identify a convenient outgroup.

A concerted effort to block the function of both copies of gene duplicates individually and in combination in various sequenced teleost genomes by mutation,

morpholino antisense oligonucleotides [92] or other strategies [93] should greatly accelerate the identification of functional noncoding sequences and conserved gene functions.

We propose that a genome-wide identification of conserved noncoding regions comparing teleost gene duplicates with their human co-orthologs, coupled with gene expression analysis to identify partitioned regulatory subfunctions and gene abrogation experiments to verify function in multiple teleosts, will contribute substantially to the identification of candidate conserved regulatory elements. This exploitation of subfunction partitioning in teleost genomes would help annotate the human genome for health, development, physiology and disease (Box 1).

References

1 Nelson, J.S. (1994) *Fishes of the World*, Wiley-Interscience
2 Ohno, S. (1970) *Evolution by Gene Duplication*, Springer Verlag
3 Ohno, C. (1973) Ancient linkage groups and frozen accidents. *Nature* 244, 259–262
4 Ekker, M. *et al*. (1992) Coordinate embryonic expression of three zebrafish *engrailed* genes. *Development* 116, 1001–1010
5 Prince, V.E. *et al*. (1998) Zebrafish *hox* genes: genomic organization and modified colinear expression patterns in the trunk. *Development* 125, 407–420
6 Van de Peer, Y. *et al*. (2002) Wanda: a database of duplicated fish genes. *Nucleic Acids Res* 30, 109–112
7 Postlethwait, J. *Fish Development and Genetics: The Zebrafish and Medaka Models* (Korzh, Z.G.a.V., ed.), World Scientific (in press)
8 Amores, A. *et al*. (1998) Zebrafish *hox* clusters and vertebrate genome evolution. *Science* 282, 1711–1714
9 Postlethwait, J. *et al*. (1998) Vertebrate genome evolution and the zebrafish gene map. *Nat. Genet.* 18, 345–349
10 Gates, M.A. *et al*. (1999) A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. *Genome Res.* 9, 334–347
11 Barbazuk, W.B. *et al*. (2000) The syntenic relationship of the zebrafish and human genomes. *Genome Res.* 10, 1351–1358
12 Postlethwait, J.H. *et al*. (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* 10, 1890–1902
13 Woods, I.G. *et al*. (2000) A comparative map of the zebrafish genome. *Genome Res.* 10, 1903–1914
14 Taylor, J.S. *et al*. (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos. Trans. R. Soc. Lond. B Biol. Sci* 356, 1661–1679
15 Postlethwait, J. *et al.* (2002) Duplication of a portion of human chromosome 20q containing Topoisomerase (Top1) and Snail genes provides evidence on genome expansion and the radiation of teleost fish. In *Aquatic Genomics: Steps Toward a Great Future* (Shimizu, N. *et al.*, eds), pp. 20–31, Springer-Verlag.
16 Van de Peer, Y. *et al*. (2002) Dealing with saturation at the amino acid level: a case study based on anciently duplicated zebrafish genes. *Gene* 295, 205–211
17 Taylor, J. *et al*. (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* 13, 382–390
18 Locke, D.P. *et al*. (2003) Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.* 13, 347–357
19 Ohno, S. (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. *Semin. Cell Dev. Biol.* 10, 517–522
20 Larhammar, D. and Risinger, C. (1994) Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*. *Mol. Phylogenet. Evol.* 3, 59–68
21 Phillips, R. and Rab, P. (2001) Chromosome evolution in the Salmonidae (Pisces): an update. *Biol. Rev. Camb. Philos. Soc.* 76, 1–25
22 David, L. *et al*. (2003) Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Mol. Biol. Evol.* 20, 1425–1434
23 Wittbrodt, J. *et al*. (1998) More genes in fish? *BioEssays* 20, 511–515
24 Meyer, A. and Schartl, M. (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* 11, 699–704
25 Gaut, B.S. and Doebley, J.F. (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. U. S. A.* 94, 6809–6814
26 Aparicio, S. *et al*. (1997) Organization of the *Fugu rubripes* Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. *Nat. Genet.* 16, 79–83
27 Vogel, G. (1998) Doubled genes may explain fish diversity. *Science* 281, 1119–1121
28 Aparicio, S. *et al*. (2000) Vertebrate evolution: recent perspectives from fish. *Trends Genet.* 16, 54–56
29 Aparicio, S. *et al*. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297, 1301–1310
30 Amores, A. *et al*. (2004) Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res.* 14, 1–10
31 Smith, S. *et al*. (2000) Identification and analysis of two snail genes in the pufferfish (*Fugu rubripes*) and mapping of human SNA to 20q. *Gene* 247, 119–128
32 Smith, S.F. *et al*. (2002) Analyses of the extent of shared synteny and conserved gene orders between the genome of *Fugu rubripes* and human 20q. *Genome Res.* 12, 776–784
33 Altschmied, J. *et al*. (2002) Subfunctionalization of duplicate *mitf* genes associated with differential degeneration of alternative exons in fish. *Genetics* 161, 259–267
34 Toramoto, T. *et al*. (2004) Multiple gene organization of pufferfish *Fugu rubripes* tropomyosin isoforms and tissue distribution of their transcripts. *Gene* 331, 41–51
35 Koopman, P. *et al*. (2004) Origin and diversity of the *Sox* transcription factor gene family: genome-wide analysis in *Fugu rubripes*. *Gene* 328, 177–186
36 Dildrop, R. and Ruther, U. (2004) Organization of Iroquois genes in fish. *Dev. Genes Evol.* 214, 26–276
37 Irwin, D.M. (2004) A second insulin gene in fish genomes. *Gen. Comp. Endocrinol.* 135, 150–158
38 Dethleffsen, K. *et al*. (2003) Insert-containing neurotrophins in teleost fish and their relationship to nerve growth factor. *Mol. Cell. Neurosci.* 24, 380–394
39 Klinger, M. *et al*. (2004) Identification of Nogo-66 receptor (NgR) and homologous genes in fish. *Mol. Biol. Evol.* 21, 76–85
40 Cardoso, J.C. *et al*. (2003) Isolation and characterisation of the corticotropin releasing factor receptor 1 (CRFR1) gene in a teleost fish, *Fugu rubripes*. *DNA Seq.* 14, 215–218
41 Yu, W.P *et al*. (2003) Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in *Fugu*. *Trends Genet.* 19, 180–183
42 Vandepoele, K. *et al*. (2004) Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1638–1643
43 Christoffels, A. *et al*. (2004) Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol. Biol. Evol.* 21, 1146–1151
44 Van de Peer, Y. *et al*. (2003) Are all fishes ancient polyploids? *J. Struct. Funct. Genomics* 3, 65–73
45 Naruse, K. *et al*. (2004) A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. *Genome Res.* 14, 820–828
46 Inoue, J.G. *et al*. (2003) Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the 'ancient fish'. *Mol. Phylogenet. Evol.* 26, 110–120
47 Chiu, C.H. *et al*. (2004) Bichir *HoxA* cluster sequence reveals surprising trends in ray-finned fish genomic evolution. *Genome Res.* 14, 11–17
48 Graham, G.J. (1995) Tandem genes and clustered genes. *J. Theor. Biol.* 175, 71–87
49 Force, A. *et al*. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545
50 Stoltzfus, A. (1999) On the possibility of constructive neutral evolution. *J. Mol. Evol.* 49, 169–181
51 Nadeau, J.H. and Sankoff, D. (1997) Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. *Genetics* 147, 1259–1266

52 Hawkins, M.B. *et al*. (2000) Identification of a third distinct estrogen receptor and reclassification of estrogen receptors in teleosts. *Proc. Natl. Acad. Sci. U. S. A.* 97, 10751–10756
53 Lister, J. *et al*. (2001) Duplicate *mitf* genes in zebrafish: complementary expression and conservation of melanogenic potential. *Dev. Biol.* 237, 333–344
54 Chiang, E.F. *et al*. (2001) Two *sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. *Dev. Biol.* 231, 149–163
55 Udono, T. *et al*. (2000) Structural organization of the human microphthalmia-associated transcription factor gene containing four alternative promoters. *Biochim. Biophys. Acta* 1491, 205–219
56 Kent, J. *et al*. (1996) A male-specific role for *SOX9* in vertebrate sex determination. *Development* 122, 2813–2822
57 Matise, M.P. and Joyner, A.L. (1997) Expression patterns of developmental control genes in normal and *Engrailed-1* mutant mouse spinal cord reveal early diversity in developing interneurons. *J. Neurosci.* 17, 7805–7816
58 Feldman, B. *et al*. (1998) Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* 395, 181–185
59 Nornes, S. *et al*. (1998) Zebrafish contains two *Pax6* genes involved in eye development. *Mech. Dev.* 77, 185–196
60 Pfeffer, P.L. *et al*. (1998) Characterization of three novel members of the zebrafish Pax2/5/8 family: Dependency of Pax5 and Pax8 expression on the Pax2.1 (noi) function. *Development* 125, 3063–3074
61 Rebagliati, M.R. *et al*. (1998) *cyclops* encodes a nodal-related factor involved in midline signaling. *Proc. Natl. Acad. Sci. U. S. A.* 95, 9932–9937
62 Oates, A.C. *et al*. (1999) Gene duplication of zebrafish *JAK2* homologs is accompanied by divergent embryonic expression patterns; only *jak2a* is expressed during erythropoiesis. *Dev. Dyn.* 215, 352–370
63 De Martino, S. *et al*. (2000) Expression of *sox11* gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. *Dev. Dyn.* 217, 279–292
64 Bruce, A.E. *et al*. (2001) Additional hox clusters in the zebrafish: divergent expression patterns belie equivalent activities of duplicate *hoxB5* genes. *Evol. Dev.* 3, 127–144
65 Chiang, E.F. *et al*. (2001) Two Cyp19 (P450 aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain. *Mol. Biol. Evol.* 18, 542–550
66 Locascio, A. *et al*. (2002) Modularity and reshuffling of Snail and Slug expression during vertebrate evolution. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16841–16846
67 McClintock, J.M. *et al*. (2002) Knockdown of duplicated zebrafish *hoxb1* genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development* 129, 2339–2354
68 Cresko, W.A. *et al*. (2003) Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. *Dev. Dyn.* 228, 480–489
69 Jozefowicz, C. *et al*. (2003) The fates of zebrafish Hox gene duplicates. *J. Struct. Funct. Genomics* 3, 185–194
70 Okubo, K. *et al*. (2003) A novel third gonadotropin-releasing hormone receptor in the medaka *Oryzias latipes*: evolutionary and functional implications. *Gene* 314, 121–131
71 Yasumoto, K. *et al*. (1998) A big gene linked to small eyes encodes multiple Mitf isoforms: many promoters make light work. *Pigment Cell Res.* 11, 329–336
72 Pöpperl, H. *et al*. (1995) Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/pbx*. *Cell* 81, 1031–1042
73 Taylor, J.S. *et al*. (2001) Genome duplication, divergent resolution and speciation. *Trends Genet.* 17, 299–301
74 Lynch, M. and Force, A. (2000) The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* 156, 590–605
75 McKinnon, J.S. *et al*. (2004) Evidence for ecology's role in speciation. *Nature* 429, 294–298
76 Kishida, M. and Callard, G.V. (2001) Distinct cytochrome P450 aromatase isoforms in zebrafish (*Danio rerio*) brain and ovary are differentially programmed and estrogen regulated during early development. *Endocrinology* 142, 740–750
77 Wittbrodt, J. *et al*. (2002) Medaka – a model organism from the Far East. *Nat. Rev. Genet.* 3, 53–64
78 Clark, M.S. *et al*. (2001) Use of the Japanese pufferfish (*Fugu rubripes*) in comparative genomics. *Mar. Biotechnol (NY)* 3, S130–S140
79 Aparicio, S. *et al*. (1995) Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 1684–1688
80 Schwartz, S. *et al*. (2003) MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences. *Nucleic Acids Res.* 31, 3518–3524
81 Mayor, C. *et al*. (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16, 1046–1047
82 Joyner, A.L. and Hanks, M. (1991) The engrailed genes: Evolution of function. *Sem. Dev. Biol.* 2, 435–445
83 Varlet, I. *et al*. (1997) Nodal expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation. *Development* 124, 1033–1044
84 Blader, P. and Strähle, U. (1998) Developmental biology: casting an eye over cyclopia. *Nature* 395, 112–113
85 Sampath, K. *et al*. (1998) Induction of the zebrafish ventral brain and floorplate requires *cyclops*/nodal signalling. *Nature* 395, 185–189
86 Dougan, S.T. *et al*. (2003) The role of the zebrafish nodal-related genes *squint* and *cyclops* in patterning of mesendoderm. *Development* 130, 1837–1851
87 Spokony, R.F. *et al*. (2002) The transcription factor Sox9 is required for cranial neural crest development in *Xenopus*. *Development* 129, 421–432
88 Mori-Akiyama, Y. *et al*. (2003) Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9360–9365
89 Yan, Y.L. *et al*. (2002) A zebrafish *sox9* gene required for cartilage morphogenesis. *Development* 129, 5065–5079
90 Bi, W. *et al*. (2001) Haploinsufficiency of *Sox9* results in defective cartilage primordia and premature skeletal mineralization. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6698–6703
91 Bi, W. (1999) *Sox9* is required for cartilage formation. *Nat. Genet.* 22, 85–89
92 Nasevicius, A. and Ekker, S.C. (2000) Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* 26, 216–220
93 Urtishak, K.A. *et al*. (2003) Targeted gene knockdown in zebrafish using negatively charged peptide nucleic acid mimics. *Dev. Dyn.* 228, 405–413
94 Kuziora, M.A. and McGinnis, W. (1988) Different transcripts of the *Drosophila* Abd-B gene correlate with distinct genetic sub-functions. *EMBO J.* 7, 3233–3244
95 Mehta, N.M. *et al*. (2003) Calcitonin for osteoporosis and bone pain. *Curr. Pharm. Des.* 9, 2659–2676