Annu. Rev. Genet. 2003. 37:123–51
doi: 10.1146/annurev.genet.37.040103.103949

First published online as a Review in Advance on June 25, 2003

PSEUDOGENES: Are They "Junk" or Functional DNA?

Evgeniy S. Balakirev[1,2] and Francisco J. Ayala[1]

[1]Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525; email: fjayala@uci.edu; [2]Institute of Marine Biology, Vladivostok 690041, Russia and Academy of Ecology, Marine Biology, and Biotechnology, Far Eastern State University, Vladivostok 690600, Russia; email: esbalak@bio.dvgu.ru

Key Words neutral evolution, evolutionary rate, DNA polymorphism, *Drosophila*, potogene

■ **Abstract** Pseudogenes have been defined as nonfunctional sequences of genomic DNA originally derived from functional genes. It is therefore assumed that all pseudogene mutations are selectively neutral and have equal probability to become fixed in the population. Rather, pseudogenes that have been suitably investigated often exhibit functional roles, such as gene expression, gene regulation, generation of genetic (antibody, antigenic, and other) diversity. Pseudogenes are involved in gene conversion or recombination with functional genes. Pseudogenes exhibit evolutionary conservation of gene sequence, reduced nucleotide variability, excess synonymous over nonsynonymous nucleotide polymorphism, and other features that are expected in genes or DNA sequences that have functional roles. We first review the *Drosophila* literature and then extend the discussion to the various functional features identified in the pseudogenes of other organisms. A pseudogene that has arisen by duplication or retroposition may, at first, not be subject to natural selection if the source gene remains functional. Mutant alleles that incorporate new functions may, nevertheless, be favored by natural selection and will have enhanced probability of becoming fixed in the population. We agree with the proposal that pseudogenes be considered as *poto*genes, i.e., DNA sequences with a *pot*entiality for becoming new genes.

CONTENTS

INTRODUCTION

Pseudogenes have been defined as nonfunctional sequences of genomic DNA that are originally derived from functional genes, but exhibit such degenerative features as premature stop codons and frameshift mutations that prevent their expression (97, 107, 132, 145, 164, 226, 231, 237). Pseudogenes are thought to arise by tandem duplication of genes, with ensuing loss of function as a result of gradual accumulation of disabling mutations (107, 132, 164, 237). "Processed" pseudogenes lack introns, and presumably arise by reverse transcription of processed mRNA, followed by integration into the genome (150, 226, 227, 231). Partially processed pseudogenes, generated by an RNA-mediated mechanism, containing or not the complete coding region, have been described in tomato (210) and human (60, 62, 194, 209). Chimeric nonfunctional genes combining parts of a functional gene and its pseudogene have been found in lemurs (108) and humans (129).

If the presence of one copy of the gene suffices for the needs of the organism, it is assumed that pseudogene mutations (disabling or not) will not be subject to purifying selection and will all have equal probability of becoming fixed in the population (85, 87, 116, 130, 131). It follows that pseudogenes will generally degenerate, owing to the rapid accumulation of recurrent mutations, and melt into the background of the surrounding DNA (e.g., 88). That process has indeed been detected in bacterial genomes (7). However, genomes of eukaryotes contain many pseudogenes that appear to have avoided full degeneration (96, 145). A large-scope analysis of pseudogene distribution based on complete genome draft sequences shows that there is less pressure to delete pseudogenes in eukaryotes than in prokaryotes (96).

We review a variety of pseudogene features in diverse organisms. We begin by reviewing the *Drosophila* literature. We then extend the discussion to other organisms, starting with cases of evolutionary conservation and proceeding to explore the various functions detected in putative pseudogenes, such as gene expression, regulation of gene expression, generation of genetic diversity (antibody, antigenic, and other), and impact on gene conversion and recombination.

We review those pseudogenes for which there is functional or evolutionary information, excluding from consideration some pseudogenes that have been characterized only by their sequence, such as tRNA (53, 197), U4 RNA (186), glutathione *S*-transferase (218, 219), and histone H2B (1).

PSEUDOGENES IN DROSOPHILA

Pseudogenes are rare in *Drosophila* (98, 161) relative to some other animals, especially vertebrates (145). DNA sequence evolution in many of the pseudogenes found in *Drosophila* manifests functional constraints, reflected in lower than expected (if the pseudogenes were not subject to selection) intraspecific variability and interspecific divergence; significant heterogeneity of nucleotide variability and divergence along the sequence; higher rate of substitution at synonymous than at nonsynonymous nucleotide positions; conservation of important functional regions; transcriptional activity; and codon bias (12, 16, 56, 110, 111, 162, 167, 208). Moreover, two *Adh* Drosophila genes originally identified as pseudogenes (70, 110) have later been considered to be novel functional genes (19, 136).

Evolutionary Conservation

Snyder et al. (203) studied a small multigene family encoding the larval cuticle proteins (LCP) of *Drosophila melanogaster*, located within a 9-kb region of the right arm of the second chromosome, at cytogenetic map position 44D. One of the five sequences was thought to be a pseudogene (ψLcp) because of several structural features and the absence of detectable transcripts. The disabling features included a 35-bp deletion eliminating the TATA box, two premature stop codons at positions 23 and 72 (162), a mutated splicing acceptor sequence, and several substitutions and indels that would prevent secretion of a functional protein (203). Another cuticle gene cluster located at 65A on the left arm of the third chromosome of *D. melanogaster* includes an intronless pseudogene, *Lcp-a*$^{\psi}$, with a 21-bp frame-shifting deletion in the signal-peptide coding region and without a consensus polyadenylation site (42).

The nucleotide polymorphism of ψLcp in *D. melanogaster* shows extensive sequence length variations that also result in premature stop codons in *D. simulans*, supporting the contention that ψLcp is not a functional gene (162). As expected in a pseudogene not subject to selection, the rates of synonymous and nonsynonymous substitutions are equal, and the overall nucleotide divergence between *D. melanogaster* and *D. simulans* is extremely high. However, the within-species nucleotide variation of ψLcp ($\pi = 0.001 \pm 0.001$) is lower than for many functional genes. The HKA test (105) reveals that the ψLcp polymorphism in *D. melanogaster* is significantly lower than expected, given the amount of divergence between *D. melanogaster* and *D. simulans*. The low level of intraspecific polymorphism of ψLcp has been attributed to background selection (162).

Currie & Sullivan (56) have characterized an intronless phosphoglyceromutase pseudogene (*Pglym87*) located on the right arm of chromosome 3 at position 87B4,5, which most likely originated from the retroposition of a *Pglym78* transcript, a gene located on the left arm of chromosome 3 at bands 78A/B. RNase protection experiments and primer extension analyses have failed to detect any transcript from *Pglym87*, although the *Pglym87* open reading frame remains

intact except for deletion of the first two codons. The structural properties suggest that *Pglym87* is a pseudogene, whereas the intact reading frame and codon bias suggest that it might be a functional gene (56).

A processed *Adh* pseudogene (*ψAdh*) has been found in *D. yakuba* and *D. teissieri* (*melanogaster* species subgroup) (110, 111). The pseudogene is located on the third chromosome, whereas the functional *Adh* gene is on the second chromosome of the species. *ψAdh* exhibits a peculiar pattern of evolution for a pseudogene, including retention of reading frame, codon bias, and a higher rate of substitution at synonymous than at nonsynonymous nucleotide positions. Long & Langley (136) have proposed that *ψAdh* is a functional gene, which they called *jingwei*. They obtained a ratio 27:4 silent:replacement polymorphism, a pattern typical of active genes; detected transcripts of the gene and suggested that it arose by retroposition from the alcohol dehydrogenase gene, followed by recruitment of additional 5′ exons and introns from an unrelated gene.

An unprocessed *Adh* pseudogene (*ψAdh*) has been detected in the *repleta* group of *Drosophila* (70; reviewed in 208). Some evidence suggests that the *ψAdh* is a pseudogene, since mutations have rendered the gene incapable of being translated into a functional alcohol dehydrogenase (208). However, the molecular evolution of *ψAdh* exhibits characteristics that are atypical of pseudogenes: (*a*) the rate of evolution is substantially slower in the exons of *ψAdh* than in the intergenic region, and only slightly faster than the rate of exons of functional *Adh* genes; (*b*) codon bias is retained in most species studied; (*c*) silent substitutions in *ψAdh* significantly exceed replacement substitutions. The Ks/Ka ratio (silent:replacement substitutions) ranges from 10 to 14, in pairwise interspecific comparisons involving sequences from seven species. These ratios greatly depart from unity, which is the value expected for a pseudogene (assuming no selective constraints), and are only slightly lower than those obtained from equivalent comparisons of the functional *Adh* genes (208). The *repleta* group *ψAdh* may be a chimeric functional gene (19).

A pseudogene has been described in the *α-esterase* gene cluster, which appears to be a bona fide pseudogene, in *D. melanogaster* (*DmαE4a-ψ*), where it has multiple inactivating mutations that are fixed in natural populations (177, 178, 185). The *α*-cluster includes 11 carboxyl/cholinesterase genes (10 active genes and 1 pseudogene) and locates within cytological region 84D3-E2 on chromosome 3R of *D. melanogaster*. The *α*-esterases have diverged substantially from one another (amino acid similarity ranging from 37% to 66%). No gene conversion or intergenic recombination has been detected between the genes and pseudogene. However, the *D. simulans* and *D. yakuba* orthologs of *DmαE4a-ψ* do not have inactivating mutations and appear to be functional (178).

The *Amylase* multigene family of *D. pseudoobscura* is located within a series of highly polymorphic inversions on the third chromosome. Four pseudogenes have been reported in three gene arrangements (ST, SC, and TL): ST *Amy3-ψ*, with a premature stop codon shortening the protein to 31.6% of its normal length (36); and TL *Amy2-ψ*, TL *Amy3-ψ*, and SC *Amy3-ψ*, with large deletions in their coding

regions (160). The divergence among the four pseudogene sequences reveals a retardation of sequence evolution in the SC and ST arrangements, whereas the rates of substitution in the two TL pseudogenes do not depart from neutral expectations. Gene conversion is most likely responsible for slowing down the divergence of SC *Amy3-ψ*, and ST *Amy3-ψ* (160). In the closely related *D. miranda*, an *Amylase* pseudogene (*Amy*3), located on the secondary sex chromosome X2, contains two large deletions (one 445 bp and the other 872 bp long) (205).

If pseudogenes are devoid of function, their pattern of nucleotide substitution is expected to reflect the pattern of spontaneous point mutations, which would make pseudogenes ideal models to study neutral evolution at the molecular level (131). Of the four *D. melanogaster Amy* pseudogenes, only two (*TL Amy*2-*ψ* and *TL Amy*3-*ψ*) seem to behave as bona fide pseudogenes, whereas in the other two, SC *Amy*3-*ψ* and ST *Amy*3-*ψ*, sequence evolution has been retarded, most likely by homogenization caused by gene conversion. Clearly, caution should be exercised when using pseudogenes as exemplars for determining patterns and rates of neutral nucleotide substitution.

Pseudogene Expression

A genomic region encompassing the *Cecropin* locus (*Cec*, located at 99E on the right arm of the third chromosome of *D. melanogaster*) includes three functional genes and two pseudogenes (*Cecψ*1 and *Cecψ*2) (126). The pseudogenes have diverged considerably from the functional *Cec* genes (only 50% of the residues remain identical), have no consensus splice signals, and contain multiple premature stop codons and deletions. However, both *Cecψ*1 and *Cecψ*2 have retained a promoter-like region with a TATA box and capping site homology (126). The two pseudogenes are also highly diverged between themselves. But nonsynonymous polymorphism is lower than synonymous polymorphism in the coding region of both pseudogenes, which suggests functional constraint on amino acid replacement changes, whereas the level of silent variation is the same (for *Cecψ*1) or higher (for *Cecψ*2) than in the functional genes. Both pseudogenes have conserved transcriptional signals and splice sites, and present an open reading frame; also, correctly spliced transcripts have been detected for both pseudogenes. *Cecψ*1 and *Cecψ*2 may be either active genes with some null alleles or young pseudogenes (167).

Regulation of Gene Expression

The X chromosome *Stellate* (*Ste*) gene of *D. melanogaster* (at 47.5 and band 12E) is homologous to a moderately repeated sequence that maps to the *Suppressor of Stellate* locus [*Su*(*Ste*)] (133) located on the long arm of the Y chromosome (93), which contains tandemly repeated pseudogene sequences with premature stop codons and frameshift mutations (17). Nevertheless, some of the tandem pseudogenes of *Su*(Ste) are transcribed (113). The Y chromosome *Su*(*Ste*) sequences regulate the activity of the *Ste* locus on the X chromosome (94, 133, 134); they repress transcription and alter the splicing of *Ste* transcripts (17, 134). The suggested

models for repression include competitive interaction between *Su*(*Ste*) and *Ste* for positively acting transcription factor(s) (113, 134); *Su*(*Ste*) antisense transcription (see below) (57); and DNA-protein or protein-protein interaction affecting *Ste* transcription (113). However, no regulatory interactions have been demonstrated for another *D. melanogaster* gene–pseudogene pair, respectively located on the X and Y chromosomes (*mst*77F gene family) (184).

The *β-esterase* Cluster of *D. melanogaster*: Expression, Evolutionary Conservation, Recombination, and Gene Conversion

Balakirev & Ayala (12) have described a "cryptic" pseudogene within the *β-esterase* gene cluster of *D. melanogaster*. The *β-esterase* gene cluster is on the left arm of chromosome 3 of *D. melanogaster*, at 68F7-69A1 in the cytogenetic map [but see (163)]. The cluster comprises two tandemly duplicated genes, first described as *Est-6* and *Est-P* (50). The coding regions are 1686 and 1691 bp long, respectively, and consist of two exons (1387 bp and 248 bp) and a small (51 bp in *Est-6* and 56 bp in *Est-P*) intron (152). The *Est-6* gene is well characterized [reviewed in (153, 175)]. The gene encodes the major β-carboxylesterase (EST-6), which is transferred by *D. melanogaster* males to females in the seminal fluid during copulation (174) and affects the female's consequent behavior and mating proclivity (90). Less information is available for *Est-P*. It was first described as a functional gene, based on several lines of evidence: transcriptional activity, intact splicing sites, no premature termination codons, and presence of initiation and termination codons (50). However, Balakirev & Ayala (12, 16) found premature stop codons within the *Est-P* coding region and some other indications suggesting that *Est-P* might be in fact a pseudogene, which they designated as *ψEst-6*. Dumancic et al. (65) showed that some alleles of *Est-P* produce a catalytically active esterase, corresponding to the previously identified EST-7 isozyme (101) and renamed the gene *Est-7*.

The comparative analysis of *ψEst-6* and *Est-6* has recently been extended to 28 strains and has included (13–16) entropy analysis, using spectral methods (43–47; reviewed in 135). Some features of *ψEst-6* indicate that it could be a pseudogene: 11 premature stop codons among 28 sequences are hardly compatible with the functionality of the encoded protein; the number of amino acid replacements is 2.9 times higher in *ψEst-6* than in *Est-6*, and some of them are drastic; nucleotide polymorphism is 2.1 times higher in *ψEst-6*; structural entropy analysis reveals significantly lower structural regularity and higher structural divergence for *ψEst-6*, in accordance with the expectations if it is a pseudogene or nonfunctional gene (16). However, as noted above, the gene can be expressed (50), and some alleles of *ψEst-6* produce a catalytically active esterase (65), although this is detected in late larvae and adults of *both* sexes, whereas the functional *Est-6* genes transcripts are found in all life stages but predominantly in adult males (50, 65), consistent with the significant role of EST-6 in male mating (90, 174). Moreover, the rate of synonymous substitutions is higher than the rate of nonsynonymous substitutions

and neutrality tests are significant (16). However, and perhaps unexpectedly, the population recombination rate is 2.6 times lower in ψ*Est-6* than in *Est-6*, so that linkage disequilibrium is more pronounced in ψ*Est-6* than in *Est-6* (81.2% versus 50.3% statistically significant associations between sites). Gene conversion occurs within both *Est-6* and ψ*Est-6*, but rarely between the two genes, and has only been detected by using the inferred protein alignment, on the basis of mutations at silent sites that occurred after the gene conversion event. The haplotype structure is dimorphic in both ψ*Est-6* and *Est-6*, each consisting of two very divergent sets of very similar alleles, but the divergent sequences of the two genes are not congruent; those of *Est-6*, but not those of ψ*Est-6*, correspond to a Slow/Fast allozyme dimorphism (16).

In conclusion, it seems that ψ*Est-6*, like other, if not most, *Drosophila* putative pseudogenes, exhibits features of both functional and nonfunctional genes (16).

EVOLUTIONARY CONSERVATION OF PSEUDOGENES

Patterns of pseudogene evolution that are inconsistent with the neutral patterns expected if the pseudogenes lacked function have been found in other organisms besides *Drosophila*. In chicken, *IglV* and *IghV* pseudogenes have been identified that include crippling mutations (182). However, only a few *IglV* and *IghV* pseudogenes contain more than one crippling mutation and, notably, very few are crippled because of stop codons or frameshift mutations. The *IglV* and *IghV* pseudogene sequences contain significantly fewer stop codons generated by point mutation than would be expected under a model of random nucleotide change. Most nucleotide substitution-generated stop codons are rescued by additional nucleotide substitutions within the same codon. The majority of insertions and deletions are in frame (182).

Furthermore, the majority of *IgV* pseudogenes have not diverged significantly from functional genes; rather, they exhibit a nonrandom distribution of variability along the complementarity-determining regions (CDR), with a peak of nucleotide and amino acid variability at the 3′ end of the pseudogenes. The framework regions (FR) appear to have been highly conserved. Thus, the chicken *IglV* and *IghV* pseudogenes exhibit evidence of selection for protein function and display all the hallmarks of expressed *IgV* genes: selection for open reading frames, diversification of the CDRs, and conservation of the FRs (182). One plausible hypothesis is that selection for open reading frames acts on these genes because they donate DNA sequence during somatic gene conversion. However, an analysis of 217 gene conversion events at the chicken *IglVI* locus reveals that pseudogenes are rarely utilized as gene conversion sequence donors; three of the pseudogenes were never used (144). Furthermore, most gene conversion events involve only a small segment of the pseudogene [mean size approximately 27 bp; (144)].

A similar state of affairs may obtain in the mouse, where a comparison of seven functional V_H genes with nine V_H pseudogenes leads to the conclusion that the *IgV*

pseudogenes may be subject to the same, or nearly the same, selection pressure as the functional genes (190). The pseudogenes have generally relatively minor defects, most frequently single nucleotide substitutions. Indels of single nucleotides leading to frameshifts or stop codons, or the loss of an initiation codon are not common. The relatively "intact" status of most immunoglobulin pseudogenes in the mouse has been noticed (22, 69, 81, 190). A single point mutation is responsible for the nonfunctionality of the human *TCRG-V10* gene (246).

An interesting example of a "preserved" pseudogene, *ψPgiC*, has been described in the wildflower genus *Clarkia* (86). Relative-rate tests show that exon nucleotides have not diverged faster in the pseudogene than in the homologous functional gene. The ratio of synonymous to replacement substitutions in the pseudogene is low, as expected if there is selection. The observed pattern cannot be accounted for by gene conversion from *PgiC1* to *ψPgiC2*. A possible explanation for the conservation of *ψPgiC2* is the occasional formation of PGIC1-PGIC2 heterodimers, so that the pseudogene may be subject to selection against mutants causing defective heterodimers (86). Experimental evidence of such deleterious subunit interactions has recently been described for Cu/Zn superoxide dismutase in *Drosophila* (158).

The conservation of pseudogene sequences retaining 90% and more homology with their functional counterparts has been reported in many organisms, such as human [steroid 21-hydroxylase pseudogene *CYP21P* (103); folate binding protein (187); laminin receptor pseudogene, *LAMRIL5* (173); macrophage stimulating protein, MSP (225)]; mouse [X-linked lymphocyte-regulated, *KLR*, unprocessed pseudogenes (79); heat-shock, *hsp25*, pseudogene (74)]; rat [*U6snRNA* pseudogene (200)]; snake [phospholipase A_2-like pseudogene (112)]; maize [*Gpa* pseudogenes (165)]; *Aspergillus nidulans* [*5SrRNA* pseudogene (25)]; and *Chlamydomonas reinhardtii* [cysteine-rich pseudogene (140)]. For additional examples, see Vanin (226), Wilde (237), and Mighell et al. (145).

The extent of similarity between the pseudogenes and their homologous functional genes is often not uniform along the sequences. In the human ferrochelatase gene and pseudogene, nucleotide identity is 82–93% in exons 2–11, but only 63% in the 5′-flanking region (234). Between a human lactate dehydrogenase gene and its pseudogene, the 5′-flanking region has twice as many differences as the coding and 3′-flanking regions (207). In salmon, there is high conservation (up to 80–90%) between the intron sequences of the growth hormone pseudogene and the functional gene, but much less homology in the coding regions (114). Similarly, in a phospholipase pseudogene from the Mojave rattlesnake, there is high conservation of the intronic regions but much less in the protein-coding regions (112). On the contrary, in the green alga *C. reinhardtii*, a cysteine-rich protein pseudogene and the functional gene are similar within coding regions, but the similarity drops dramatically in the intron and in the 5′- and 3′-noncoding regions. The first and second exon sequences show a 91% and 74% similarity, respectively, whereas the 5′- and 3′-flanking regions are only 51% and 41% similar. Similarity between the intron sequences is even lower, 40%, and is clustered toward the exon-intron boundaries (140).

PSEUDOGENE TRANSCRIPTION AND EXPRESSION

The original definition of pseudogenes implied that they should be transcriptionally and translationally silent and hence not subject to selection (131). However, nonfunctional as well as functional transcripts of pseudogenes have been described in many organisms. The human examples are numerous and include interferon pseudogene ψ*LeIFNE* [(84); *IFNAP22* (63)]; glyceraldehydes-3-phosphate dehydrogenase pseudogene ψ*Gapdh* (9, 224); glucocerebrosidase pseudogene *psGBA* (204); dopamine D5 pseudogene ψ*DRD5-1* (148); dopamine $D_{1\beta}$ pseudogene (232); complement factor H-related protein 1 pseudogene *H36-2* (201); DNA topoisomerase 1 pseudogene *TOP1* (248); serotonin 5-hydroxytryptamine receptor ψ*5HT*$_{1D\alpha}$ (18); glutamine synthetase Ψ*GS* (39); neuropeptide Y1-like receptor ψ*NPY Y1-like* (181); steroid 21-hydrolase pseudogene *CYP21P* or *CYP21A* (67); type I hair keratin Ψ*hHaA*, which is transcribed in human (180), whereas its orthologs are also expressed in chimpanzee and gorilla (238); tumor repressor Ψ*PTEN* (75); 5-HT7 receptor Ψ*5-HT7* (154); translationally controlled tumor protein pseudogene *TPT1*, transcribed in all tissues investigated, but with widely variable expression in different human tissues, whereas its orthologs are fully expressed in the rabbit (215); and myosin pseudogene *MYO15BP* (24).

Pseudogene transcripts have been detected in other mammals, including mouse glyceraldehyde-3-phosphate dehydrogenase ψ*Gapdh* (77, 224); bovine aromatase pseudogene *Cyp19*ψ (76) and cytochrome b_5 pseudogene (55); and ribonuclease pseudogene *BS RNAase* in sheep (52). Other animal examples include the silk moth chorion locus, high-cysteine protein pseudogene ψ*HcB.12/13* (72). Plant examples are *rps19* pseudogene in *Oenothera berteriana* (191), *rps14* pseudogene (10) and myrosinase pseudogene *TGG3* (245) in *Arabidopsis thaliana*, potato ribosomal protein ψ*rps14* (166), and liverwort NADH dehydrogenase pseudogene (211). Other examples are erythrocyte binding protein Ψ*EBA165* in *Plasmodium falciparum* (222) and yeast adenine phosphoribosyltransferase pseudogene *APT2* (2).

Human pseudogene translation activity has been shown in vivo (33, 142) as well as in vitro (146; for mouse, see 27) or, simply, suggested (18, 232). The human *glucocerebrosidase* pseudogene *psGBA* promoter has some activity when attached to a reporter gene (169). This pseudogene is transcriptionally active in vivo and the amount of pseudogene-derived mRNA is sometimes comparable to the amount of mRNA derived from the active gene, indicating that pseudogene transcription occurs at a high level (204). However, there is considerable subject-to-subject variation in the amount of pseudogene transcript relative to the functional gene, which could be due to polymorphism in the promoter regions of the pseudogene or to polymorphisms affecting the stability of pseudogene-derived mRNA (204). Interestingly, the pseudogene transcripts can be detected in tissues not known to express the functional counterpart gene (154), or in only some of the tissues where the functional homolog is specifically expressed (245). This situation is reminiscent of ψ*Est-6* in *Drosophila melanogaster*, which is transcribed mostly

in late larvae and adults of each sex, whereas the functional *Est-6* gene transcripts are detected in all life stages but predominate in adult males (50, 65).

REGULATION OF GENE EXPRESSION

McCarrey & Riggs (141) have suggested a possible role for pseudogenes in development as a source of intracellular inhibitors. They hypothesized that pseudogenes could be the source of the antisense RNA that hybridizes with the sense RNA from the determinator genes, thereby blocking their expression. Pseudogenes, or portions thereof, may be transcribed from the opposite strand relative to their functional counterparts, which would make them a source of antisense RNA. If the inhibitors act at the protein level, then they should be identifiable, for example, as proteins that modify the electrophoretic migration of determinator proteins. One prediction derived from this hypothesis has been confirmed, namely that the sequence of duplicated genes that are transcribed in opposite directions should be conserved whenever the sense-antisense interaction is advantageous (141).

There are numerous examples in various organisms of unusually high level of pseudogene conservation, exceeding 90% of the sequence, as pointed out above. Moreover, pseudogenes are prevalent in the genomes of multicellular organisms and scarce in those of unicellular organisms, such as bacteria and yeast (97, 128), which is consistent with the hypothesis of McCarrey & Riggs (141). Korneev et al. (122) have shown that a *nitric oxide synthase* pseudogene (*pseudo-NOS*) and its paralogous functional gene (*nNOS*) are co-expressed in identifiable neurons of the mollusk *Lymnaea stagnalis*. The *pseudo-NOS* transcript includes a region with significant antisense homology to the *nNOS* mRNA. The antisense region of the *pseudo-NOS* RNA specifically suppresses the synthesis of the *nNOS* protein (122). Thus the *pseudo-NOS* transcript acts as an antisense regulator of *nNOS* protein synthesis.

There has been much speculation and some evidence that pseudogenes may have regulatory roles for the genes from which they have been derived (72, 73, 106, 248). Healy et al. (100) have shown that 3′ sequences that lie within the ψ*Est-6* pseudogene transcription unit of *D. melanogaster* contain elements that modulate the expression of *Est-6*, which obviously implies some regulatory function for ψ*Est-6*. Sequence similarity exists in a 390-bp block within the 609-bp 3′-flanking region of *Est-P* (*D. melanogaster*) and the 3′-flanking sequence of *Est-5A* (*D. pseudoobscura*) (28, 50). In particular, a segment of 110 bp within this region shows 76% sequence similarity between the *Est-P* and *Est-5A*, which sharply contrasts with 20% or lower similarity in the 5′ flanking region between the same two genes or between *Est-6* and *Est-5B* (*D. melanogaster* versus *D. pseudoobscura*) (28). Obvious reductions in intraspecific variation as well as interspecific divergence also occur in this region between *D. melanogaster* and *D. simulans* (16).

Troyanovsky & Leube (223) have described an interesting example of gene/pseudogene cooperation in human *cytokeratin 17* expression. A detailed examination of *cytokeratin* transcription regulation using gene/pseudogene chimeric

constructs has identified specific promoter/enhancer elements that are inactive by themselves but can interact to induce strong transcriptional activity of reporter genes. The process includes the interaction between the proximal region of the inactive *cytokeratin* pseudogene promoter and the distal upstream region of the actively transcribed *cytokeratin* gene. Troyanovsky & Leube (223) conclude that *cis* elements in the proximal 5′-upstream region of the pseudogene promoter can cooperate with distal enhancer elements of the functional gene to induce strong transcriptional activity in transfected HeLa cells. In mice, the *Makorin1-p1* pseudogene regulates the messenger-RNA stability of its homologous coding gene (*Makorin1*) by competitive interaction, either at the RNA or DNA level (104a). Similar mechanisms have been previously suggested for regulatory gene-pseudogene interaction (113, 134).

PSEUDOGENES AS RESERVOIRS FOR GENERATING GENETIC DIVERSITY

Antibody Diversity

A functional role for pseudogenes in the immune response has been proposed for human (4, 155, 228, 242), chicken (20, 143, 144, 171, 172, 189), rabbit (71, 118), and other vertebrates. Immunoglobulin gene diversity is generated by somatic gene conversion events in which sequences derived from alleles or paralogous genes generate new gene sequences. Gene conversion may also occur between genes and pseudogenes, so that selection may contribute to conserving the sequence and other functional characteristics of pseudogenes. In chicken, the process has been shown to occur in the germ line as well as somatically during antibody diversification (20). Together with the observations that the immunoglobulin pseudogenes are highly conserved, have open reading frames, and retain canonical structure sequence patterns, these observations support the hypothesis that the immunoglobulin pseudogenes resemble a functional multigene family maintained by selection for its functional role in generating somatic antibody diversity (143).

Antigenic Variation

Pseudogenes act as a reservoir of sequences that recombine with their functional paralogous genes and thus generate antigenic diversity. Gene conversion events have been reported in several bacterial pathogens as a mechanism for generating sequence diversity in expressed antigenic genes. In the bacterium *Borrelia hermsii*, a *vmp* pseudogene serves as a source of antigenic variation during relapsing fever (170). Bacteria of the genus *Borrelia* generate antigenic diversity of the *vmp/vls* coat proteins through recombination from a tandem array of silent partial pseudogene cassettes into a telomeric expression site on a linear plasmid (244). Antigenic variation is generated in *Anaplasma marginale* (a member of ehrlichial genogroup II) by recombination of pseudogenes into the functional expression site (30, 31).

Pseudogenes are also a source of genetic diversity in gonococci. These bacteria generate pilin variants that allow evasion of the host immune system. In this process, a silent, partial pilin gene sequence is introduced into the expression locus to form a complete pilin gene (91). Pilin phase variation can also occur by deletions between direct repeats internal to the expressed pilin gene or between a pilin gene and an adjacent silent pseudogene (104).

Antigenic variation of surface proteins in *Mycoplasma* (151) is achieved by recombination between multiple pseudogenes and the expressed *vlhA* gene; the multiple pseudogenes serve as a repertoire for the creation of *vlhA* sequence diversity, probably by site-specific recombination, with recombination between the single complete *vlhA* gene and one of the multiple partial copies creating new *vlhA* gene variants (151). Subsequent gene conversion, drawing on the shorter pseudogenes, may introduce further variability into this new variant. Thus, the expressed genes are likely to be chimeric, with sequences derived from several overlapping pseudogenes (151). The discrete boundaries of sequence similarity observed for the pseudogenes strongly suggest a site-specific mechanism of gene conversion (151).

In *Trypanosoma cruzi* antigenic diversity is generated by segmental gene conversion (3). Recombination between *vsg* genes and pseudogenes has been suggested as a mechanism by which antigenic diversity is generated in *Trypanosoma equiperdum* (217), as well as in the *T. cruzi* genome, by recombination between surface protein pseudogenes (212, 213).

Other Gene Systems

The human olfactory receptor (OR) pseudogenes may be important for the generation and maintenance of receptor diversity (82). Intensive intergenic gene conversion has been revealed for this multigene family that leads to segment shuffling in the odorant binding site, an evolutionary process reminiscent of somatic combinatorial diversification in the immune system (196). Although OR pseudogenes have lost full coding function, they are apparently under new evolutionary constraints: OR pseudogenes adopt noncoding functions as CpG islands (82), enhancers (37), and matrix attachment regions (80).

RECOMBINATION AND SEQUENCE POLYMORPHISM

Recombination is a key enabling mechanism in the continuing evolution of life on earth. It is widely believed to have originated in bacteria or their progenitors as a DNA repair process. The biochemistry of recombination has reinforced a long-established link between recombination and repair (54, 125, 179). Actively transcribed yeast genes show increased recombination rates when compared to identical, nontranscribed genes (216).

Disruption of homologous sequences by indels or nucleotide polymorphism can significantly reduce recombination frequencies (32, 83, 230). Recombination

within the maize *bz1* gene is inhibited when one allele with a transposon insertion is paired with a second allele lacking the insertion (64). A similar outcome was observed at the maize *a1* locus (241). In *D. melanogaster* the observed reduction of recombination in *ψEst-6* and the absence of gene conversion between *Est-6* and *ψEst-6* (16) may be due in some strains to indels, such as the insertion of the *mdg-3* retrotransposon within the intron of *ψEst-6* detected in some strains (16, 78).

The recombination machinery may be sensitive even to a single nucleotide mismatch. Individual nucleotide substitutions have been shown to affect recombination in yeast (26, 195), bacteria (198, 199), and mammalian cells (137). Strong inhibition of intrachromosomal recombination, sometimes by as much as 100- to 1000-fold, due to small sequence divergence, has been observed in the mouse (60, 229), mammalian cells (66, 243), yeast (11, 48, 58, 59, 95, 195), and bacteria (49, 139, 168). Inhibition is largely mediated by the mismatch repair system, which is thought to scan hybrid DNA and abort recombination when too many mismatches are detected (41, 58, 59, 61, 147, 195). Conversion tract lengths for spontaneous ectopic events are apparently reduced by sequence divergence, an effect that can also be explained by hybrid DNA rejection (48, 95).

As noted earlier, the recombination rate is lower in *ψEst-6* of *D. melanogaster* than in the *Est-6* gene. Moreover, no (or little) gene conversion occurs between *Est-6* and *ψEst-6* (16). The nucleotide sequences of the coding regions of *Est-6* and *ψEst-6* show 64% similarity (50). This level of nucleotide divergence probably does not satisfy the homology requirements for efficient intergenic conversion and represents sufficient sequence divergence to disrupt conversion between the genes. However, interlocus gene conversion occurs between *Est-5A* and *Est-5B* of *D. pseudoobscura* that are orthologous to *Est-6* and *ψEst-6*, respectively, of *D. melanogaster*, but exhibit much higher similarity, 82.5%, between the coding regions (28, 117).

Recombination is not random within the human β-globin gene cluster (115, 233). There is a "hot spot" of recombination between the δ and β genes (115). This phenomenon could be caused by specific sequences promoting recombination. The 5′-flanking region of the gene is known to have a chi-like sequence (5′-GTCGGTGG-3′), which is presumably a promoter of recombination (40). Chi-like sequences are involved in recombination in bacteriophage λ (202), murine immunoglobulin genes (124), human and murine class I and class II genes in the major histocompatibility complex (239), and in human *CYP21* genes (6, 129). The sequence GGXGGX has been implicated to function as a hot spot for gene conversion events in an early chorion multigene family (102, 123). The conversion events between Cβ1 and Cβ2 genes of the T-cell antigen receptor β-chain locus in murine (183) and rabbit (138) are limited to exon 1. Specific recognition sequences that promote conversion were suggested to explain the conversion pattern observed (183).

Other examples of specific sequences promoting recombination are suggested in the vertebrate *Major Histocompatibility Complex* (*MHC*-) *DRB* genes. Every second intron of all investigated *MHC-DRB* genes contains a defined stretch of tandemly repeated simple DNA sequences (5, 176, 192), the basic structure of

which is $(gt)_n(ga)_m$. These simple sequences have been found in proximity to presumed hot spots of recombination in the mouse *MHC* (119) and the human *HLA-DQ* region (188). This fact has led to the suggestion that a causal relation exists between the simple repeated DNA structures and the recombination hot spots in the *MHC* [reviewed in (188, 206)]. It is interesting that unprecedented evolutionary stability of simple repeats in the expressed mammalian *MHC-DRB* genes occurs in some specific genome locations (5, 8, 29, 89, 176).

In mammalian *DRB* pseudogenes, the simple repeat stretch seems to gradually reduce its characteristic pattern in the evolutionary course. Particularly, the *HLA-DRBX* pseudogene (127) harbors only a short $(ga)_6(ca)_8$ stretch, hence lacking the basic $(gt)_n(ga)_m$ structure. The simple repeat structure of the sheep *Ovar-DRB3* pseudogene is highly disintegrated and in the *Ovar-DRB4* pseudogene only three copies of each dinucleotide remained detectable. Simple repeat sequences in the human pseudogene *DRB2* also show signs of shortening and disintegration (176).

Defective DNA recombination signals in the T-cell receptor A joining-region pseudogenes have been identified in mice (121), primates (214), and human (120). Absence of recombination signal sequences or crippling mutations in these signal sequences have been revealed in chicken *IglV* and *IghV* pseudogenes (171, 172). Bliskovskii et al. (23) have described the structure of the human *son* processed pseudogene, which has a 96% homology with the *son* functional gene. Despite the high sequence homology, the *son* pseudogene lacks five monomers of the perfect tandem repeat area, which are present in the homologous *son* gene region and which, as suggested, are associated with the initiation of recombination processes (51, 109, 221).

In several yeast genes, the sequence signals initiating recombination often occur within the promoter but not within the gene itself (38, 68, 157, 236, 240). Disruption of the *HIS4*, *GAL10*, and *ARG4* promoter regions inhibit recombination and gene conversion within these loci (149, 157, 216, 235). Even partial deletions of the yeast *ARG4* and *HIS4* promoter regions inhibit recombination within these genes (149, 235). A role for promoter sequences in initiating recombination has been identified in maize *a1* (241) and *b1* (156) genes.

In *D. melanogaster*, the promoter of ψ*Est-6* is limited to the 193 bp encompassing the intergenic region between *Est-6* and ψ*Est-6* (50), and it may be that the observed reduction in the rate of recombination of ψ*Est-6* is associated with the promoter's truncation. The sequences promoting recombination could also be eroded within ψ*Est-6* owing to stochastic accumulation of mutations, as in cases of the *HLA-DRB* (127, 176) and *son* (23) pseudogenes.

CONCLUSIONS

We have described in *Drosophila* and other organisms, including humans, pseudogene features that would be unexpected if pseudogenes were nonfunctional sequences of genomic DNA ("junk" DNA). An extensive and fast-increasing

literature does not justify a sharp division between genes and pseudogenes that would place pseudogenes in the class of genomic "junk" DNA that lacks function and is not subject to natural selection. Pseudogenes are often extremely conserved and transcriptionally active. Moreover, there are direct indications that pseudogenes can be functional by being involved in the regulation of gene expression and in generating genetic diversity. Eukaryote genomes contain many pseudogenes that have avoided strong degeneration because they are important components of the genome (34, 35, 96). Pseudogenes may lose some specific functions but retain others, and even acquire new ones, which may not be simply recognizable.

Pseudogenes may recover the full original function of the genes from which they derive. Pseudogene function has been restored in vitro by mutagenesis (247), transfection (193), or in vivo by site-specific (or intermolecular) recombination (21, 170). Possible restoration of function by gene conversion has been proposed for the ribonuclease pseudogene in artiodactyl evolution (220). In the chlorella virus PBCV-1, a nonfunctional, nontranscribed cytosine DNA methyltransferase pseudogene differs by eight amino acids from a functional counterpart, but a single amino acid change or deletion by site-directed mutagenesis restores the pseudogene's activity (247). Restoration by site directed mutagenesis has been obtained in the yeast *Schizosaccharomyces pombe* (159). A strain of *E. coli* deleted for the *lacZ* gene has been shown to give rise to spontaneous lactose-utilizing mutants as a result of the activation of the *ebg* pseudogene [reviewed in (92)].

How pervasive are "functional" pseudogenes? Many pseudogenes have been identified in all sorts of organisms on the grounds that they are duplicated genes that exhibit stop codons or other disabling mutations in their DNA sequences, so that they cannot have the full function of the original genes from which derived. In many of these cases, however, it remains unknown, because it has not been investigated, whether the pseudogenes, described only on the basis of DNA sequences, may have acquired regulatory or other functions, or play a role in generating genetic variability. There seems to be the case that some functionality has been discovered in all cases, or nearly, whenever this possibility has been pursued with suitable investigations. One may well conclude that most pseudogenes retain or acquire some functionality and, thus, that it may not be appropriate to define pseudogenes as nonfunctional sequences of genomic DNA originally derived from functional genes, or as "genes that are no longer expressed but bear sequence similarity to active genes" (99, p. 114). Rather, pseudogenes might be defined as DNA sequences derived by duplication or retroposition from functional genes that are often subject to natural selection and therefore retain much of the original sequence and structure because they have acquired new regulatory or other functions, or may serve as reservoirs of genetic variability.

Much remains to be learned about the population dynamics of pseudogenes. This much is clear: It is unwarranted to assume that "all mutations occurring in pseudogenes are selectively neutral and become fixed in the population with equal probability" (87, p. 124). Once a pseudogene appears in a population, presumably in a single genome, by duplication or retroposition, the pseudogene may at first

follow neutral population dynamics and become extinct or not. A pseudogene may rapidly become fixed if it is closely linked to a gene experiencing a selective sweep, or slowly increase in frequency by neutral drift. However, a mutant that provides a functional role for a pseudogene may be favored by natural selection and, thus, have a higher probability of becoming fixed than its nonfunctional alleles (137a,b). It may be appropriate to include pseudogenes within the class named by Brosius & Gould (34) as "potogenes." These authors pointed out that the products of gene duplication, including those that become pseudogenes, may eventually acquire distinctive functions, and thus might be called *poto*genes to call attention to their *pot*entiality for becoming new genes.

ACKNOWLEDGMENTS

This work is supported by NIH grant GM42397 to F. J. Ayala.

The *Annual Review of Genetics* is online at http://genet.annualreviews.org

LITERATURE CITED

1. Akhmanova A, Hennig W. 1998. *Drosophila melanogaster* histone H2B retropseudogene is inserted into a region rich in transposable elements. *Genome* 41: 396–401
2. Alfonzo JD, Crother TR, Guetsova ML, Daignan-Fornier B, Taylor MW. 1999. *APT1*, but not *APT2*, codes for a functional adenine phosphoribosyltransferase in *Saccharomyces cerevisiae*. *J. Bacteriol.* 181:347–52
3. Allen CL, Kelly JM. 2001. *Trypanosoma cruzi*: mucin pseudogenes organized in a tandem array. *Exp. Parasitol.* 97:173–77
4. Almagro JC, Domínguez-Martinez V, Lara-Ochoa F, Vargas-Madrazo E. 1995. Structural repertoire in human V_L pseudogenes of immunoglobulins: comparison with functional germline genes and amino acid sequences. *Immunogenetics* 43:92–96
5. Ammer H, Schwaiger F-W, Kammerbauer C, Gomolka M, Arriens A, et al. 1992. Exonic polymorphism versus intronic simple repeat hypervariability in *MHC-DRB* genes. *Immunogenetics* 35: 332–40
6. Amor M, Parker KL, Globerman H, New MI, White PC. 1988. Mutation in the *CYP21B* gene (Ile172→Asn) causes steroid 21-hydroxylase deficiency. *Proc. Natl. Acad. Sci. USA* 85:1600–4
7. Andersson JO, Andersson SG. 2001. Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes. *Mol. Biol. Evol.* 18:829–39
8. Andersson L, Sigurdardottir S, Borsch C, Gustafsson K. 1991. Evolution of MHC polymorphism: extensive sharing of polymorphic sequence motifs between human and bovine DRB alleles. *Immunogenetics* 33:188–93
9. Arcari P, Martinelli R, Salvatore F. 1984. The complete sequence of a full length cDNA for human liver glyceraldehyde-3-phosphate dehydrogenase: evidence for multiple mRNA species. *Nucleic Acids Res.* 12:9179–89
10. Aubert D, Bisanz-Seyer C, Herzog M. 1992. Mitochondrial *rps14* is a transcribed and edited pseudogene in *Arabidopsis thaliana*. *Plant. Mol. Biol.* 20: 1169–74
11. Bailis AM, Rothstein R. 1990. A defect

in mismatch repair in *Saccharomyces cerevisiae* stimulates ectopic recombination between homeologous genes by an excision repair dependent process. *Genetics* 126:535–47

12. Balakirev ES, Ayala FJ. 1996. Is esterase-P encoded by a cryptic pseudogene in *Drosophila melanogaster*? *Genetics* 144:1511–18
13. Balakirev ES, Ayala FJ. 2003. Nucleotide variation of the *Est-6* gene region in natural populations of *Drosophila melanogaster*. *Genetics.* In press
14. Balakirev ES, Balakirev EI, Ayala FJ. 2002. Molecular evolution of the *Est-6* gene in *Drosophila melanogaster*: contrasting patterns of DNA variability in adjacent functional regions. *Gene* 288:167–77
15. Balakirev ES, Balakirev EI, Rodriguez-Trelles F, Ayala FJ. 1999. Molecular evolution of two linked genes, *Est-6* and *Sod*, in *Drosophila melanogaster*. *Genetics* 153:1357–69
16. Balakirev ES, Chechetkin VR, Lobzin VV, Ayala FJ. 2003. DNA polymorphism in the β*-esterase* gene cluster of *Drosophila melanogaster*. *Genetics* 164: 533–44
17. Balakireva MD, Shevelyov YY, Nurminsky DI, Livak KJ, Gvozdev VA. 1992. Structural organization and diversification of *Y*-linked sequences comprising *Su*(*Ste*) genes in *Drosophila melanogaster*. *Nucleic Acids Res.* 20: 3731–36
18. Bard JA, Nawoschik SP, O'Dowd BF, George SR, Branchek TA, Weinshank RL. 1995. The human serotonin 5-hydroxytryptamine$_{1D}$ receptor pseudogene is transcribed. *Gene* 153:295–96
19. Begun D. 1997. Origin and evolution of a new gene descended from *alcohol dehydrogenase* in Drosophila. *Genetics* 145:375–82
20. Benatar T, Ratcliffe MJH. 1993. Polymorphism of the functional immunoglobulin variable region genes in the chicken by exchange of sequence with donor pseudogenes. *Eur. J. Immunol.* 23:2448–53
21. Benevolenskaya EV, Kogan GL, Tulin AV, Philipp D, Gvozdev VA. 1997. Segmented gene conversion as a mechanism of correction of 18S rRNA pseudogene located outside of rDNA cluster in *D. melanogaster*. *J. Mol. Evol.* 44:646–51
22. Blankenstein T, Bonhomme F, Krawinkel U. 1987. Evolution of pseudogenes in the immunoglobulin VH-gene family of the mouse. *Immunogenetics* 26:237–48
23. Bliskovskii VV, Kirillov AV, Spirin KS, Zakharev VM, Chumakov IM. 1993. *Son* pseudogene does not contain five repeated elements of the area of perfect tandem repeats present in the homologous *son* gene sequence. *Mol. Biol.* 27:61–68
24. Boger ET, Sellers JR, Friedman TB. 2001. Human myosin XVBP is a transcribed pseudogene. *J. Muscle Res. Cell Motil.* 22:477–83
25. Borsuk P, Gniadkowski M, Bartnik E, Stepien PP. 1988. Unusual evolutionary conservation of 5S rRNA pseudogenes in *Aspergillus nidulans*: similarity of the DNA sequence associated with the pseudogenes with the mouse immunoglobulin switch region. *J. Mol. Evol.* 28:125–30
26. Borts RH, Haber JE. 1987. Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science* 237:1459–65
27. Boschan C, Borchert A, Ufer C, Thiele BJ, Kuhn H. 2001. Discovery of a functional retrotransposon of the murine phospholipid hydroperoxide glutathione peroxidase: chromosomal localization and tissue-specific expression pattern. *Genomics* 79:387–94
28. Brady JP, Richmond RC. 1992. An evolutionary model for the duplication and divergence of esterase genes in Drosophila. *J. Mol. Evol.* 34:506–21

29. Braunstein NS, Germain RN. 1986. The mouse E-beta-2 gene: a class II MHC beta gene with limited intraspecies polymorphism and an unusual pattern of transcription. *EMBO J.* 5:2469–76
30. Brayton KA, Knowles DP, McGuire TC, Palmer GH. 2001. Efficient use of a small genome to generate antigenic diversity in tick-borne ehrlichial pathogens. *Proc. Natl. Acad. Sci. USA* 98:4130–35
31. Brayton KA, Palmer GH, Lundgren A, Yi J, Barbet AF. 2002. Antigenic variation of *Anaplasma marginale msp2* occurs by combinatorial gene conversion. *Mol. Microbiol.* 43:1151–59
32. Brenner DA, Smigocki AC, Camerini-Otero RD. 1985. Effect of insertions, deletions, and double-strand breaks on homologous recombination in mouse L cells. *Mol. Cell. Biol.* 5:684–91
33. Bristow J, Gitelman SE, Tee MK, Staels B, Miller WL. 1993. Abundant adrenal-specific transcription of the human P450c21A "pseudogene". *J. Biol. Chem.* 268:12919–24
34. Brosius J, Gould SJ. 1992. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc. Natl. Acad. Sci. USA* 89:10706–10
35. Brosius J, Tiedge H. 1996. Reverse transcriptase: mediator of genomic plasticity. *Virus Genes* 11:163–79
36. Brown CJ, Aquadro CF, Anderson WW. 1990. DNA sequence evolution of the amylase multigene family in *Drosophila pseudoobscura*. *Genetics* 126:131–38
37. Buettner JA, Glusman G, Ben-Arie N, Ramos P, Lancet D, Evans GA. 1998. Organization and evolution of olfactory receptor genes on human chromosome 11. *Genomics* 53:56–68
38. Cao L, Alani E, Kleckner N. 1990. A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. *Cell* 61:1089–101
39. Chakrabarti R, McCracken JB, Chakrabarti D, Souba WW. 1995. Detection of a functional promoter/enhancer in an intronless human gene encoding a glutamine synthase-like enzyme. *Gene* 153:163–99
40. Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH. 1984. Nonuniform recombination within the human β-globin gene cluster. *Am. J. Hum. Genet.* 36:1239–58
41. Chambers SR, Hunter N, Louis EJ, Borts RH. 1996. The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol. Cell. Biol.* 16:6110–20
42. Charles J-P, Chihara C, Nejad S, Riddiford LM. 1997. A cluster of cuticle protein genes of *Drosophila melanogaster* at 65A: sequence, structure and evolution. *Genetics* 147:1213–24
43. Chechetkin VR, Knizhnikova LA, Turygin AY. 1994. Three-quasiperiodicity, mutual correlations, ordering and long-range modulations in genomic nucleotide sequences for viruses. *J. Biomol. Struct. Dyn.* 12:271–99
44. Chechetkin VR, Lobzin VV. 1996. Levels of ordering in coding and non-coding regions of DNA sequences. *Phys. Lett. A* 222:354–60
45. Chechetkin VR, Lobzin VV. 1998. Study of correlations in segmented DNA sequences: application to structural coupling between exons and introns. *J. Theor. Biol.* 190:69–83
46. Chechetkin VR, Turygin AY. 1994. On the spectral criteria of disorder in non-periodic sequences: application to inflation models, symbolic dynamics and DNA sequences. *J. Phys. A: Math. Gen.* 27:4875–98
47. Chechetkin VR, Turygin AY. 1996. Study of correlations in DNA sequences. *J. Theor. Biol.* 178:205–17
48. Chen W, Jinks-Robertson S. 1999. The role of mismatch repair machinery in

regulating mitotic and meiotic recombination between diverged sequences in yeast. *Genetics* 151:1299–313

49. Claverys JP, Lacks SA. 1986. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol. Rev.* 50:133–65
50. Collet C, Nielsen KM, Russell RJ, Karl M, Oakeshott JG, Richmond RC. 1990. Molecular analysis of duplicated esterase genes in *Drosophila melanogaster*. *Mol. Biol. Evol.* 7:9–28
51. Collick A, Jeffreys AJ. 1990. Detection of a novel minisatellite-specific DNA-binding protein. *Nucleic Acids Res.* 18: 625–29
52. Confalone E, Beintema JJ, Sasso MP, Carsana A, Palmieri M, et al. 1995. Molecular evolution of genes encoding ribonucleases in ruminant species. *J. Mol. Evol.* 41:850–58
53. Cooley L, Schaack J, Burke DJ, Thomas B, Söll D. 1984. Transcription factor binding is limited by the 5′-flanking regions of a *Drosophila* $tRNA^{His}$ gene and a $tRNA^{His}$ pseudogene. *Mol. Cell. Biol.* 4:2714–22
54. Cox MM. 1993. Relating biochemistry to biology: How the recombinational repair function of the recA system is manifested in its molecular properties. *BioEssays* 15:617–23
55. Cristiano RJ, Giordano SJ, Steggles AW. 1993. The isolation and characterization of the bovine cytochrome b_5 gene, and a transcribed pseudogene. *Genomics* 17:348–54
56. Currie PD, Sullivan DT. 1994. Structure, expression and duplication of genes which encode phosphoglyceromutase of *Drosophila melanogaster*. *Genetics* 138:353–63
57. Danilevskaya ON, Kurenova EV, Pavlova MN, Bebehov DV, Link AJ, et al. 1991. He-T family DNA sequences in the *Y* chromosome of *Drosophila melanogaster* share homology with the *X*-linked *Stellate* genes. *Chromosome* 100:118–24
58. Datta A, Adjiri A, New L, Grouse GF, Jinks-Robertson S. 1996. Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 176:1085–93
59. Datta A, Hendrix M, Lipsitch M, Jinks-Robertson S. 1997. Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl. Acad. Sci. USA* 94:9757–62
60. de Coo R, Buddiger P, Smeets H, van Kessel AG, Morgan-Hughes J, et al. 1995. Molecular cloning and characterization of the active human mitochondrial NADH: ubiquinone oxidoreductase 24-kDa gene (*NDUFV2*) and its pseudogene. *Genomics* 26:461–66
61. De Wind N, Dekker M, Berns A, Radman M, Riele HT. 1995. Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination and predisposition to cancer. *Cell* 82:321–30
62. Dhawan P, Yang E, Kumar A, Mehta KD. 1998. Genetic complexity of the human geranylgeranyltransferase I beta-subunit gene: a multigene family of pseudogenes derived from mis-spliced transcripts. *Gene* 210:9–15
63. Díaz M, Pomykala HM, Bohlander SK, Maltepe E, Malik K, et al. 1994. Structure of the human type-I interferon gene cluster determined from a YAC clone contig. *Genomics* 22:540–52
64. Dooner HK. 1986. Genetic fine structure of the *bronze* locus in maize. *Genetics* 113:1021–36
65. Dumancic MM, Oakeshott JG, Russell RJ, Healy MJ. 1997. Characterization of the *EstP* protein in *Drosophila melanogaster* and its conservation in Drosophilids. *Biochem. Genet.* 35:251–71
66. Elliott B, Richardson C, Winderbaum J, Nickoloff JA, Jasin M. 1998. Gene conversion tracts in mammalian cells from

double-strand break repair. *Mol. Cell. Biol.* 18:93–101
67. Endoh A, Yang L, Hornsby PJ. 1998. CYP21 pseudogene transcripts are much less abundant than those from the active gene in normal human adrenocortical cells under various conditions in culture. *Mol. Cell. Endocrinol.* 137:13–19
68. Fan Q, Xu F, Petes T. 1995. Meiosis-specific double-strand breaks at the *HIS4* recombination hot spot in the yeast *Saccharomyces cerevisiae* control in *cis* and *trans*. *Mol. Cell. Biol.* 15:1679–88
69. Ferguson SE, Rudikoff S, Osborne BA. 1988. Interaction and sequence diversity among T15 V_H genes in CBA/J mice. *J. Exp. Med.* 168:1339–49
70. Fisher JA, Maniatis T. 1985. Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. *Nucleic Acids Res.* 13:6899–917
71. Fitts MG, Metzger DW. 1990. Identification of rabbit genomic *Ig-V_H* pseudogenes that could serve as donor sequences for latent allotype expression. *J. Immunol.* 145:2713–17
72. Fotaki ME, Iatrou K. 1988. Identification of a transcriptionally active pseudogene in the chorion locus of the silkmoth *Bombyx mori*. Regional sequence conservation and biological function. *J. Mol. Biol.* 203:849–60
73. Fotaki ME, Iatrou K. 1993. Silk moth chorion pseudogenes: hallmarks of genomic evolution by sequence duplication and gene conversion. *J. Mol. Evol.* 37:211–20
74. Fröhli E, Aoyama A, Klemenz R. 1993. Cloning of the mouse *hsp25* gene and an extremely conserved *hsp25* pseudogene. *Gene* 128:273–77
75. Fujii GH, Morimoto AM, Berson AE, Bolen JB. 1990. Transcriptional analysis of the PTEN/MMAC1 pseudogene, ΨPTEN. *Oncogene* 18:1765–69
76. Fürbass R, Vanselow J. 1995. An aromatase pseudogene is transcribed in the bovine placenta. *Gene* 154:287–91
77. Galland F, Stefanova M, Pirisi V, Birnbaum D. 1990. Characterization of a murine glyceraldehyde-3-phosphate dehydrogenase pseudogene. *Biochimie* 72: 759–62
78. Game AY, Oakeshott JG. 1990. Associations between restriction site polymorphism and enzyme activity variation for esterase 6 in *Drosophila melanogaster*. *Genetics* 126:1021–31
79. Garchon H-J, Loh E, Ho WY, Amar L, Avner P, Davis MM. 1989. The XLR sequence family: dispersion on the X and Y chromosomes of a large set of closely related sequences, most of which are pseudogenes. *Nucleic Acids Res.* 17:9871–88
80. Gimelbrant AA, McClintock TS. 1997. A nuclear matrix attachment region is highly homologous to a conserved domain of olfactory receptors. *J. Mol. Neurosci.* 9:61–63
81. Givol D, Zakut R, Effron K, Rechavi G, Ram D, Cohen JB. 1981. Diversity of germ-line immunoglobulin VH genes. *Nature* 292:426–30
82. Glusman G, Sosinsky A, Ben-Asher E, Avidan N, Sonkin D, et al. 2000. Sequence, structure, and evolution of a complete human olfactory receptor gene cluster. *Genomics* 63:227–45
83. Godwin AR, Liskay RM. 1994. The effects of insertions on mammalian intrachromosomal recombination. *Genetics* 136:607–17
84. Goeddel DV, Leung DW, Dull TJ, Gross M, Lawn RM, et al. 1981. The structure of eight distinct cloned human leukocyte interferon cDNAs. *Nature* 290:20–26
85. Gojobori T, Li W-H, Graur D. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. *J. Mol. Evol.* 18:360–69
86. Gottlieb LD, Ford VS. 1997. A recently silenced, duplicate *PgiC* locus in *Clarkia*. *Mol. Biol. Evol.* 14:125–32
87. Graur D, Li W-H. 2000. *Fundamentals of Molecular Evolution*, p. 124. Sunderland, MA: Sinauer. 2nd ed.

88. Graur D, Shuali Y, Li W-H. 1989. Deletions in processed pseudogenes accumulate faster in rodents than in humans. *J. Mol. Evol.* 28:279–85
89. Groenen MAM, van der Poel JJ, Dijkhof RJM, Giphart RJ. 1990. The nucleotide sequence of bovine MHC class II DQB and DRB genes. *Immunogenetics* 31:37–44
90. Gromko MH, Gilbert DF, Richmond RC. 1984. Sperm transfer and use in the multiple mating system of Drosophila. In *Sperm Competition and the Evolution of Animal Mating Systems*, ed. RL Smith, pp. 371–426. New York: Academic
91. Haas R, Veit S, Meyer TF. 1992. Silent pilin genes of *Neisseria gonorrhoeae* MS11 and the occurrence of related hypervariant sequences among other gonococcal isolates. *Mol. Microbiol.* 6:197–208
92. Hall BG. 1990. Directed evolution of a bacterial operon. *BioEssays* 12:551–58
93. Hardy RW, Lindsley DL, Livak KJ, Lewis B, Siversten AL, et al. 1984. Cytogenetic analysis of a segment of the *Y* chromosome of *Drosophila melanogaster*. *Genetics* 107:591–610
94. Hardy RW, Tokuyasu KT, Lindsley DL. 1981. Analysis of spermatogenesis in *Drosophila melanogaster* bearing deletions for *Y* chromosome fertility genes. *Chromosoma* 83:593–617
95. Harris S, Rudnicki KS, Haber JE. 1993. Gene conversions and crossing over during homologous and homeologous ectopic recombination in *Saccharomyces cerevisiae*. *Genetics* 135:5–16
96. Harrison PM, Gerstein M. 2002. Studying genomes through the aeons: protein families, pseudogenes and proteome evolution. *J. Mol. Biol.* 318:1155–74
97. Harrison PM, Kumar A, Lan N, Echols N, Snyder M, Gerstein M. 2002. A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. *J. Mol. Biol.* 316:409–19
98. Harrison PM, Milburn D, Zhang Z, Bertone P, Gerstein M. 2003. Identification of pseudogenes in the *Drosophila melanogaster* genome. *Nucleic Acids Res.* 31:1033–37
99. Hartl DL, Clark AG. 1989. *Principles of Population Genetics.* Sunderland, MA: Sinauer. 2nd ed.
100. Healy MJ, Dumancic MM, Cao A, Oakeshott JG. 1996. Localization of sequences regulating ancestral and acquired sites of esterase 6 activity in *Drosophila melanogaster*. *Mol. Biol. Evol.* 13:784–97
101. Healy MJ, Dumancic MM, Oakeshott JG. 1991. Biochemical and physiological studies of soluble esterases from *Drosophila melanogaster*. *Biochem. Genet.* 29:365–88
102. Hibner BL, Burke WD, Eickbush TH. 1991. Sequence identity in an early chorion multigene family is the result of localized gene conversion. *Genetics* 128:595–606
103. Higashi Y, Yoshioka H, Yamane M, Gotoh O, Fujii-Kuriyama Y. 1986. Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. *Proc. Natl. Acad. Sci. USA* 83:2841–45
104. Hill SA, Morrison SG, Swanson J. 1990. The role of direct oligonucleotide repeats in gonococcal pilin gene variation. *Mol. Microbiol.* 4:1341–52
104a. Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, et al. 2003. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature 423:91–96
105. Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–59
106. Inouye M. 1988. Antisense RNA: its functions and applications in gene regulation—a review. *Gene* 72:25–34
107. Jacq C, Miller JR, Brownlee GG. 1977.

A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* 12:109–20
108. Jeffreys AJ, Barrie PA, Harris S, Fawcett DM, Nugent ZJ, Boyd AC. 1982. Isolation and sequence analysis of a hybrid δ-globin pseudogene from the brown lemur. *J. Mol. Biol.* 156:487–95
109. Jeffreys AJ, Wilson V, Thein SL. 1985. Hypervariable "minisatellite" regions in human DNA. *Nature* 314:67–73
110. Jeffs PS, Ashburner M. 1991. Processed pseudogenes in Drosophila. *Proc. R. Soc. London Ser. B* 244:151–59
111. Jeffs PS, Holmes EC, Ashburner M. 1994. The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. *Mol. Biol. Evol.* 11:287–304
112. John TR, Smith JJ, Kaiser II. 1996. A phospholipase A_2-like pseudogene retaining the highly conserved introns of Mojave toxin and other snake venom group II PLA_2s, but having different exons. *DNA Cell Biol.* 15:661–68
113. Kalmykova AI, Dobritsa AA, Gvozdev VA. 1998. *Su*(*Ste*) diverged tandem repeats in a *Y* chromosome of *Drosophila melanogaster* are transcribed and variously processed. *Genetics* 148:243–49
114. Kavsan VM, Koval AP, Palamarchuk AJ. 1994. A growth hormone pseudogene in the salmon genome. *Gene* 141:301–2
115. Kazazian HH, Chakravarti A, Orkin SH, Antonarakis SE. 1983. DNA polymorphism in the human β-globin gene cluster. In *Evolution of Genes and Proteins*, ed. M Nei, R Koehn, pp. 137–46. Sunderland, MA: Sinauer
116. Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–20
117. King LM. 1998. The role of gene conversion in determining sequence variation and divergence in the *Est-5* gene family in *Drosophila pseudoobscura*. *Genetics* 148:305–15
118. Knight KL. 1992. Restricted V_H usage and generation of antibody diversity in rabbit. *Annu. Rev. Immunol.* 10:593–616
119. Kobori JA, Winoto A, McNicholas J, Hood L. 1984. Molecular characterization of the recombination region of six murine major histocompatibility complex (MHC) I-region recombinants. *Mol. Cell. Immunol.* 1:125–31
120. Koop BF, Rowen L, Wang K, Kuo CL, Seto D, et al. 1994. The human T-cell receptor *TCRAC/TCRAD* (C alpha/C delta) region: organization, sequence and evolution of 97.6 kb DNA. *Genomics* 19:478–93
121. Koop BF, Wilson RK, Wang K, Vernooij B, Zaller D, et al. 1992. Organization, structure and function of 95 kb of DNA spanning the murine T-cell receptor C alpha/C delta region. *Genomics* 13:1209–30
122. Korneev SA, Park J-H, O'Shea M. 1999. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J. Neurosci.* 19:7711–20
123. Kravariti L, Lecanidou R, Rodakis GC. 1995. Sequence analysis of a small early chorion gene subfamily interspersed within the late gene locus in *Bombyx mori*. *J. Mol. Evol.* 41:24–33
124. Krawinkel U, Zoebelein G, Bruggemann M, Radbruch A, Rajewsky K. 1983. Recombination between antibody heavy chain variable-region genes: evidence for gene conversion. *Proc. Natl. Acad. Sci. USA* 80:4997–5001
125. Kuzminov A. 1996. *Recombinational Repair of DNA Damage*. Georgetown, TX: R. G. Landes
126. Kylsten P, Samakovlis C, Hultmark D. 1990. The cecropin locus in Drosophila; a compact gene cluster involved in the response to infection. *EMBO J.* 9:217–24

127. Larhammar D, Servenius B, Rask L, Peterson PA. 1985. Characterization of an HLA DR-beta pseudogene. *Proc. Natl. Acad. Sci. USA* 82:1475–79
128. Lawrence JG, Hendrix RW, Casjens S. 2001. Where are the pseudogenes in bacterial genomes? *Trends Microbiol.* 9: 535–40
129. Lee H-H, Niu D-M, Lin R-W, Chan P, Lin C-Y. 2002. Structural analysis of the chimeric *CYP21P/CYP21* gene in steroid 21-hydroxylase deficiency. *J. Hum. Genet.* 47:517–22
130. Li W-H. 1983. Evolution of duplicate genes and pseudogenes. In *Evolution of Genes and Proteins*, ed. M Nei, RK Koehn, pp. 14–37. Sunderland, MA: Sinauer
131. Li W-H, Gojobori T, Nei M. 1981. Pseudogenes as a paradigm of neutral evolution. *Nature* 292:237–39
132. Little PF. 1982. Globin pseudogenes. *Cell* 28:683–84
133. Livak KJ. 1984. Organization and mapping of a sequence on the *Drosophila melanogaster X* and *Y* chromosomes that is transcribed during spermatogenesis. *Genetics* 107:611–34
134. Livak KJ. 1990. Detailed structure of the *Drosophila melanogaster Stellate* genes and their transcripts. *Genetics* 124:303–16
135. Lobzin VV, Chechetkin VR. 2000. Order and correlations in genomic DNA sequences. The spectral approach. *Usp. Fiz. Nauk* 170:57–81 In Russian. English transl. *Phys. Usp.* 43:55–78
136. Long MY, Langley CH. 1993. Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* 260:91–95
137. Lukacsovich T, Waldman AS. 1999. Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. *Genetics* 151:1559–68
137a. Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–55
137b. Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* 159:1789–804
138. Marche PN, Kindt TJ. 1986. A variable region gene subfamily encoding T cell receptor beta-chains is selectively conserved among mammals. *J. Immunol.* 137:1729–34
139. Matic I, Rayssiguier C, Radman M. 1995. Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* 80:507–15
140. Matters GL, Goodenough UW. 1992. A gene/pseudogene tandem duplication encodes a cysteine-rich protein expressed during zygote development in *Chlamydomonas reinhardtii*. *Mol. Gen. Genet.* 232:81–88
141. McCarrey JR, Riggs AD. 1986. Determinator-inhibitor pairs as a mechanism for threshold setting in development: a possible function for pseudogenes. *Proc. Natl. Acad. Sci. USA* 83:679–83
142. McCarrey JR, Thomas K. 1987. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature* 326:501–5
143. McCormack WH, Hurley EA, Thompson CB. 1993. Germ line maintenance of the pseudogene donor pool for somatic immunoglobulin gene conversion in chickens. *Mol. Cell. Biol.* 13:821–30
144. McCormack WH, Thompson CB. 1990. *IgL* variable gene conversion display pseudogene donor preference and 5′ to 3′ polarity. *Genes Dev.* 4:548–58
145. Mighell AJ, Smith NR, Robinson PA, Markham AF. 2000. Vertebrate pseudogenes. *FEBS Lett.* 468:109–14
146. Misra-Press A, Cooke NE, Liebhaber SA. 1994. Complex alternative splicing partially inactivates the human chorionic somatomammotropin-like (*hCS-L*) gene. *J. Biol. Chem.* 269:23220–29

147. Negritto MT, Wu X, Kuo T, Chu S, Bailis AM. 1997. Influence of DNA sequence identity on efficiency of targeted gene replacement. *Mol. Cell. Biol.* 17:278–86
148. Nguyen T, Sunahara R, Marchese A, Van Tol HHM, Seeman P, O'Dowd BF. 1991. Transcription of a human dopamine pseudogene. *Biochem. Biophys. Res. Commun.* 181:16–21
149. Nicolas A, Treco D, Schultes NP, Szostak JW. 1989. An initiation site for meiotic gene conversion in the yeast *Saccharomyces cerevisiae*. *Nature* 338:35–39
150. Nishioka Y, Leder A, Leder P. 1980. Unusual α-globin-like gene that has cleanly lost both globin intervening sequences. *Proc. Natl. Acad. Sci. USA* 77:2806–9
151. Noormohammadi AH, Markham PF, Kanci A, Whithear KG. Browning GF. 2000. A novel mechanism for control of antigenic variation in the haemagglutinin gene family of *Mycoplasma synoviae*. *Mol. Microbiol.* 35:911–23
152. Oakeshott JG, Collet C, Phillis R, Nielsen KM, Russell RJ, et al. 1987. Molecular cloning and characterization of esterase 6, a serine hydrolase from Drosophila. *Proc. Natl. Acad. Sci. USA* 84:3359–63
153. Oakeshott JG, van Papenrecht EA, Boyce TM, Healy MJ, Russell RJ. 1993. Evolutionary genetics of Drosophila esterases. *Genetica* 90:239–68
154. Olsen MA, Schechter LE. 1999. Cloning, mRNA localization and evolutionary conservation of a human 5-HT7 receptor pseudogene. *Gene* 227:63–69
155. Pascual V, Capra D. 1991. Human immunoglobulin heavy-chain variable region genes: organization, polymorphism, and expression. *Adv. Immunol.* 49:1–74
156. Patterson GI, Kubo KM, Shroyer T, Chandler VL. 1995. Sequences required for paramutation of the maize *b* gene map to a region containing the promoter and upstream sequences. *Genetics* 140: 1389–406
157. Petes TD, Malone RE, Symington LS. 1991. Recombination in yeast. In *The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis, and Energetics*, ed. JR Broach, JR Pringle, EW Jones, 1:407–521. Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press
158. Phillips JP, Tainer JA, Getzoff ED, Boulianne GL, Kirby K, Hilliker AJ. 1995. Subunit-destabilizing mutations in Drosophila copper/zinc superoxide dismutase: neuropathology and a model of dimer disequilibrium. *Proc. Natl. Acad. Sci. USA* 92:8574–78
159. Pinarbasi E, Elliott J, Hornby DP. 1996. Activation of a yeast pseudo DNA methyltransferase by deletion of a single amino acid. *J. Mol. Biol.* 257:804–13
160. Popadić A, Norman RA, Doanet WW, Anderson WW. 1996. The evolutionary history of the amylase multigene family in *Drosophila pseudoobscura*. *Mol. Biol. Evol.* 13:883–88
161. Powell JR. 1997. *Progress and Prospects in Evolutionary Biology. The Drosophila Model*. Oxford/New York: Oxford Univ. Press
162. Pritchard JK, Schaeffer SW. 1997. Polymorphism and divergence at a *Drosophila* pseudogene locus. *Genetics* 147:199–208
163. Procunier WS, Smith JJ, Richmond RC. 1991. Physical mapping of the *esterase-6* locus of *Drosophila melanogaster*. *Genetica* 84:203–8
164. Proudfoot N. 1980. Pseudogenes. *Nature* 286:840–41
165. Quigley F, Brinkmann H, Martin WF, Cerff R. 1989. Strong functional GC-pressure in a light-regulated maize gene encoding subunit GAPA of chloroplast glyceraldehyde-3-phosphate dehydrogenase: implications for the evolution of GAPA pseudogenes. *J. Mol. Evol.* 29:412–21
166. Quiñones V, Zanlungo S, Moenne A, Gómez I. 1996. The *rpl5-rps14-cob*

gene arrangement in *Solanum tuberosum*: *rps14* is a transcribed and unedited pseudogene. *Plant Mol. Biol.* 31:937–43

167. Ramos-Onsins S, Aguadé M. 1998. Molecular evolution of the *Cecropin* multigene family in Drosophila: functional genes *vs*. pseudogenes. *Genetics* 150:157–71
168. Rayssiguier C, Thaler DS, Radman M. 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch repair mutants. *Nature* 342:396–401
169. Reiner O, Horowitz M. 1988. Differential expression of the human glucocerebrosidase-coding gene. *Gene* 73:469–78
170. Restrepo BI, Carter CJ, Barbour AG. 1994. Activation of a *vmp* pseudogene in *Borrelia hermsii*: an alternate mechanism of antigenic variation during relapsing fever. *Mol. Microbiol.* 13:287–99
171. Reynaud CA, Anquez V, Dahan A, Weill JC. 1987. A hyperconversion mechanism generates the chicken preimmune light chain repertoire. *Cell* 48:379–88
172. Reynaud CA, Dahan A, Anquez V, Weill JC. 1989. Somatic hyperconversion diversifies the single VH gene of the chicken with a high incidence in the D region. *Cell* 59:171–83
173. Richardson MP, Braybrook C, Tham M, Moore GE, Stanier P. 1998. Molecular cloning and characterization of a highly conserved human 67-kDa laminin receptor pseudogene mapping to Xq21.3. *Gene* 206:145–50
174. Richmond RC, Gilbert DG, Sheehan KB, Gromko MH, Butterworth FM. 1980. Esterase 6 and reproduction in *Drosophila melanogaster*. *Science* 207:1483–85
175. Richmond RC, Nielsen KM, Brady JP, Snella EM. 1990. Physiology, biochemistry and molecular biology of the *Est-6* locus in *Drosophila melanogaster*. In *Ecological and Evolutionary Genetics of Drosophila*, ed. JSF Barker, WT Starmer, RJ MacIntyre, pp. 273–92. New York: Plenum
176. Riess O, Kammerbauer C, Roewer L, Steimle V, Andreas A, et al. 1990. Hypervariability of intronic simple $(gt)_n(ga)_m$ repeats in *HLA-DRB* genes. *Immunogenetics* 32:110–16
177. Robin GCQ, Medveczky KM, Russell RJ, Oakeshott JG. 1996. Duplication and divergence of the genes of the alpha-esterase cluster of *Drosophila melanogaster*. *J. Mol. Evol.* 43:241–52
178. Robin GCQ, Russell RJ, Cutler DJ, Oakeshott JG. 2000. The evolution of an α-esterase pseudogene inactivated in the *Drosophila melanogaster* lineage. *Mol. Biol. Evol.* 17:563–75
179. Roca AI, Cox MM. 1997. RecA protein: structure, function, and role in recombinational DNA repair. *Prog. Nucleic Acids Res. Mol. Biol.* 56:129–23
180. Rogers MA, Winter H, Wolf C, Heck M, Schweizer J. 1998. Characterization of a 190–kilobase pair domain of human type I hair keratin genes. *J. Biol. Chem.* 273:26683–91
181. Rose PM, Lynch JS, Frazier ST, Fisher SM, Chung W, et al. 1997. Molecular genetic analysis of a human neuropeptide Y receptor. The human homolog of the murine "Y5" receptor may be a pseudogene. *J. Biol. Chem.* 272:3622–27
182. Rothenfluh HS, Blanden RV, Steele EJ. 1995. Evolution of V genes: DNA sequence structure of functional germline genes and pseudogenes. *Immunogenetics* 42:159–71
183. Rudikoff S, Fitch WM, Heller M. 1992. Exon-specific gene correction (conversion) during short evolutionary periods: homogenization in a two-gene family encoding the β-chain constant region of the T-lymphocyte antigen receptor. *Mol. Biol. Evol.* 9:14–26
184. Russel SR, Kaiser K. 1993. *Drosophila melanogaster* male germ line-specific transcripts with autosomal and *Y*-linked genes. *Genetics* 134:293–308
185. Russell RJ, Robin GC, Kostakos P, Newcomb RD, Boyce TM, et al.

1995. Molecular cloning of an α esterase gene cluster on chromosome 3R of *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* 26:235–47

186. Saba JA, Busch H, Wright D, Reddy R. 1986. Isolation and characterization of two putative full-length *Drosophila* U4 small nuclear RNA genes. *J. Biol. Chem.* 261:8750–53
187. Sadasivan E, Cedeno M, Rothenberg SP. 1992. Genomic organization of the gene and a related pseudogene for a human folate binding protein. *Biochim. Biophys. Acta* 1131:91–94
188. Satyanarayana K, Strominger JL. 1992. NA sequences near a meiotic recombinational breakpoint within the human HLA-DQ region. *Immunogenetics* 35: 235–40
189. Sayegh CE, Drury G, Ratcliffe JH. 1999. Efficient antibody diversification by gene conversion *in vivo* in the absence of selection for V(D)J-encoded determinants. *EMBO J.* 18:6319–28
190. Schiff C, Milili M, Fougereau M. 1985. Functional and pseudogenes are similarly organized and may equally contribute to the extensive antibody diversity of the *IgVHII* family. *EMBO J.* 4:1225–30
191. Schuster W, Brennicke A. 1991. RNA editing makes mistakes in plant mitochondria: editing loses sense in transcripts of a *rps19* pseudogene and in creating stop codons in *coxl* and *rps3* mRNAs of *Oenothera*. *Nucleic Acids Res.* 19:6923–28
192. Schwaiger F-W, Weyers E, Buitkamp J, Ede AJ, Crawford A, Epplen JT. 1994. Interdependent *MHC-DRB* exon-plus-intron evolution in artiodactyls. *Mol. Biol. Evol.* 11:239–49
193. Seiser C, Beck G, Wintersberger E. 1990. The processed pseudogene of mouse thymidine kinase is active after transfection. *FEBS Lett.* 270:123–26
194. Selig S, Bruno S, Scharf JM, Wang CH, Vitale E, et al. 1995. Expressed cadherin pseudogenes are localized to the critical region of the spinal muscular atrophy gene. *Proc. Natl. Acad. Sci. USA* 92:3702–6
195. Selva EM, New L, Crouse GF, Lahue RS. 1995. Mismatch correction acts as a barrier to homologous recombination in *Saccharomyces cerevisiae*. *Genetics* 139:1175–88
196. Sharon D, Glusman G, Pilpel Y, Khen M, Gruetzner F, et al. 1999. Primate evolution of an olfactory receptor cluster: Diversification by gene conversion and recent emergence of pseudogenes. *Genomics* 61:24–36
197. Sharp S, DeFranco D, Silberklang M, Hosbach HA, Schmidt T, et al. 1981. The initiator tRNA genes in *Drosophila melanogaster*: evidence for a tRNA pseudogene. *Nucleic Acids Res.* 9:5867–82
198. Shen P, Huang HV. 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* 112:441–57
199. Shen P, Huang HV. 1989. Effect of base pair mismatches on recombination via the RecBCD pathway. *Mol. Gen. Genet.* 218:358–60
200. Sibbald PR, Blencowe BJ. 1990. A completely conserved rat U6 snRNA pseudogene coding sequence is sandwiched between a cytochrome *c* retropseudogene and a LINE-like sequence. *Nucleic Acids. Res.* 18:1063
201. Skerka C, Horstmann RD, Zipfel PF. 1991. Molecular cloning of a human serum protein structurally related to complement factor H. *J. Biol. Chem.* 266:12015–20
202. Smith GR, Kunes SM, Schultz DW, Taylor A, Triman KL. 1981. Structure of Chi hotspots of generalized recombination. *Cell* 24:429–36
203. Snyder M, Hunkapiller M, Yuen D, Silvert D, Fristrom J, Davidson N. 1982. Cuticle protein genes of Drosophila: structure, organization and evolution of

four clustered genes. *Cell* 29:1027–40

204. Sorge J, Gross E, West C, Beutler E. 1990. High level transcription of the glucocerebrosidase pseudogene in normal subjects and patients with Gaucher disease. *J. Clin. Invest.* 86:1137–41
205. Steinemann S, Steinemann M. 1999. The *Amylase* gene cluster on the evolving sex chromosomes of *Drosophila miranda*. *Genetics* 151:151–61
206. Steinmetz M, Stephan D, Fischer-Lindahl K. 1986. Gene organisation and recombinatorinal hotspots in the murine major histocompatibility complex. *Cell* 44:895–904
207. Sudo K, Maekawa M, Luedemann MM, Deaven LL, Li SSL. 1990. Human lactate dehydrogenase-B processed pseudogene: nucleotide sequence analysis and assignment to the X-chromosome. *Biochem. Biophys. Res. Commun.* 171: 67–74
208. Sullivan DT, Starmer WT, Curtiss SW, Menotti-Raymond M, Yum J. 1994. Unusual molecular evolution of an *Adh* pseudogene in Drosophila. *Mol. Biol. Evol.* 11:443–58
209. Suzuki E, Lowry J, Sonoda G, Testa JR, Walsh K, 1996. Structures and chromosome locations of the human *MEF2A* gene and a pseudogene *MEF2AP*. *Cytogenet. Cell. Genet.* 73:244–49
210. Szkukálek A, Solymosy F. 1994. Molecular characterization of two tomato U6 RNA pseudogenes generated by RNA-mediated mechanisms. *Plant Sci.* 99:183–87
211. Takemura M, Nozato N, Oda K, Kobayashi Y, Fukuzawa H, Ohyama K. 1995. Active transcription of the pseudogene for subunit 7 of the NADH dehydrogenase in *Marchantia polymorpha* mitochondria. *Mol. Gen. Genet.* 247: 565–70
212. Takle GB, O'Connor J, Young AJ, Cross GA. 1992. Sequence homology and absence of mRNA defines a possible pseudogene member of the *Trypanosoma cruzi* gp85/sialidase multigene family. *Mol. Biochem. Parasitol.* 56:117–28
213. Taylor MC, Muhia DK, Baker DA, Mondragon A, Schaap P, Kelly JM. 1999. *Trypanosoma cruzi* adenylyl cyclase is encoded by a complex multigene family. *Mol. Biochem. Parasitol.* 104:205–17
214. Thiel C, Otting N, Bontrop RE, Lanchbury JS. 1995. Generation and reactivation of T-cell receptor A joining region pseudogenes in primates. *Immunogenetics* 43:57–62
215. Thiele H, Berger M, Skalweit A, Thiele BJ. 2000. Expression of the gene and processed pseudogenes encoding the human and rabbit translationally controlled tumour protein (TCTP). *Eur. J. Biochem.* 267:5473–81
216. Thomas BJ, Rothstein R. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* 56:619–30
217. Thon G, Baltz T, Eisen H. 1989. Antigenic diversity by the recombination of pseudogenes. *Genes Dev.* 3:1247–54
218. Toba G, Aigaki T. 2000. Disruption of the *Microsomal glutathione S-transferase-like* gene reduces life span of *Drosophila melanogaster*. *Gene* 253:179–87
219. Toung Y-PS, Hsieh TS, Chen-Pei DT. 1993. The glutathion S-transferase D genes: a divergently organized, intronless gene family in *Drosophila melanogaster*. *J. Biol. Chem.* 268:9737–46
220. Trabesinger-Ruef N, Jermann T, Zankel T, Durrant B, Frank G, Benner SA. 1996. Pseudogenes in ribonuclease evolution: a source of new biomacromolecular function? *FEBS Lett.* 382:319–22
221. Treco D, Arnheim N. 1986. The evolutionarily conserved repetitive sequence $d(TG\ AG)_n$ promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Mol. Cell. Biol.* 6:3934–47
222. Triglia T, Thompson JK, Cowman AF. 2001. An EBA175 homologue which is transcribed but not translated in

erythrocytic stages of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 116: 55–63

223. Troyanovsky SM, Leube RE. 1994. Activation of the silent human cytokeratin 17 pseudogene-promoter region by cryptic enhancer elements of the cytokeratin 17 gene. *Eur. J. Biochem.* 223:61–69
224. Tso J, Sun XH, Kao TH, Reece K, Wu R. 1985. Isolation and characterization of rat and human glyceraldehyde-3-phosphate dehydrogenase cDNAs: genomic complexity and molecular evolution of the gene. *Nucleic Acids Res.* 13:2485–502
225. van der Drift P, Chan A, Zehetner G, Westerveld A, Versteeg R. 1999. Multiple MSP pseudogenes in a local repeat cluster on 1p36.2: an expanding genomic graveyard? *Genomics* 62:74–81
226. Vanin E. 1985. Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* 19:253–72
227. Vanin EF, Goldberg GI, Tucker PW, Smithies O. 1980. A mouse α-globin-related pseudogene lacking intervening sequences. *Nature* 268:222–26
228. Vargas-Madrazo E, Almagro JC, Lara-Ochoa F. 1995. Structural repertoire in V_H pseudogenes of immunoglobulins: comparison with human germline genes and human amino acid sequences. *J. Mol. Biol.* 246:74–81
229. Waldman AS, Liskay RM. 1987. Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. *Proc. Natl. Acad. Sci. USA* 84:5340–44
230. Waldman AS, Liskay RM. 1988. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.* 8:5350–57
231. Weiner AM, Deininger PL, Efstratiadis A. 1986. Nonviral retroposons: gene, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55:631–61
232. Weinshank RL, Adham N, Macchi M, Olsen MA, Branchek TA, Hartig PR. 1991. Molecular cloning and characterization of a high affinity dopamine receptor ($D1_\beta$) and its pseudogene. *J. Biol. Chem.* 266:22427–35
233. Weir BS, Hill WG. 1986. Nonuniform recombination within the human β-globin gene cluster. *Am. J. Hum. Genet.* 38:776–78
234. Whitcombe DM, Albertson DG, Cox TM. 1994. Molecular analysis of functional and nonfunctional genes for human ferrochelatase: isolation and characterization of a FECH pseudogene and its sublocalization on chromosome 3. *Genomics* 20:482–86
235. White MA, Detloff P, Strand M, Petes TD. 1992. A promoter deletion reduces the rate of mitotic, but not meiotic recombination at the *HIS4* locus in yeast. *Curr. Genet.* 21:109–16
236. White MA, Dominska M, Petes TD. 1993. Transcription factors are required for the meiotic recombination hotspot at the *HIS4* locus in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 90:6621–25
237. Wilde CD. 1986. Pseudogenes. *CRC Crit. Rev. Biochem.* 19:323–52
238. Winter H, Langbein L, Krawczak M, Cooper DN, Jave-Suarez LF, et al. 2001. Human type I hair keratin pseudogene ψ*hHaA* has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the *Pan-Homo* divergence. *Hum. Genet.* 108:37–42
239. Wu S, Saunders TL, Bach FH. 1986. Polymorphism of human Ia antigens generated by reciprocal intergenic exchange between two DR β loci. *Nature* 324:676–79
240. Wu T-C, Lichten M. 1994. Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science* 263:515–18
241. Xu X, Hsia A-P, Zhang L, Nikolau BJ,

Schnable PS. 1995. Meiotic recombination break points resolve at high rates at the 5′ end of a maize coding sequence. *Plant Cell* 7:2151–61

242. Wysocki LJ, Gefter ML. 1989. Gene conversion and the generation of antibody diversity. *Annu. Rev. Biochem.* 58:509–31
243. Yang D, Waldman AS. 1997. Fine-resolution analysis of products of intrachromosomal homeologous recombination in mammalian cells. *Mol. Cell. Biol.* 17: 3614–28
244. Zhang J-R, Hardham JM, Barbour AG, Norris SJ. 1997. Antigenic variation in Lyme disease Borrelia by promiscuous recombination of VMP-like sequence cassettes. *Cell* 89:275–85
245. Zhang J, Pontoppidan B, Xue J, Rask L, Meijer J. 2002. The third myrosinase gene TGG3 in *Arabidopsis thaliana* is a pseudogene specifically expressed in stamen and petal. *Physiol. Plant.* 115:25–34
246. Zhang X-M, Cathala G, Soua Z, Lefranc P-M, Huck S. 1996. The human T-cell receptor gamma variable pseudogene *V10* is a distinctive marker of human speciation. *Immunogenetics* 43:196–203
247. Zhang Y, Nelson M, Van Etten JL. 1992. A single amino acid change restores DNA cytosine methyltransferase activity in a cloned Chlorella virus pseudogene. *Nucleic Acids Res.* 20:1637–42
248. Zhou B-S, Beidler DR, Cheng Y-C. 1992. Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene. *Cancer Res.* 52:4280–85

Annual Review of Genetics
Volume 37, 2003

CONTENTS