Original Articles

Eukaryotic Operon Genes Can Define Highly Conserved Syntenies

(eukaryotic operon / conserved synteny)

Z. TRACHTULEC

Department of Mammalian Molecular Genetics and Centre for Integrated Genomics, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Abstract. The synteny conservation of the members of eukaryotic operons was investigated by mapping their orthologues in Drosophila, human, and other eukaryotes. While the homologues of the operon members are generally not linked, some examples of highly conserved syntenies were found. The most significant synteny involves two members of one *C. elegans* operon, encoding fibrillarin and ribosomal protein S16. Their homologues are linked in human, mouse, Drosophila, *Anopheles gambiae, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Plasmodium falciparum*, and *Guillardia theta*, but not in five other genomes. The distances between the genes are larger than in the nematode, suggesting the prevalence of intrachromosomal rearrangements.

The pre-mRNAs of at least 13% of *C. elegans* genes are transcribed in operons (Zorio et al., 1994; Blumenthal et al., 2002). Unlike bacteria, *C. elegans* trans-splices its polycistronic transcipts to monocistronic, and the proportion of more genes of the same operon annotated for the same biological process is much lower (von Mering and Bork, 2002).

Operons also exist in other nematodes (Evans et al., 1997), but in other animals, only a few multicistronic transcripts arising from nuclear genes were detected (Schulz et al., 1990; Lee, 1991; Andrews et al., 1996; Brogna and Ashburner, 1997; Ibnsouda et al., 1998; Reiss et al., 1999; Gray and Nicholls, 2000; Liu et al., 2000). These animal operon genes are not trans-spliced. One of the human operons was inspected in diverse animal species and found to be conserved not just in vertebrates, but also in *C. elegans* and Drosophila (Gray and Nicholls, 2000).

Received January 30, 2004. Accepted February 2, 2004.

This work was supported by grant No. 204/01/0997 from the Czech Science Foundation.

Corresponding author: Zdenek Trachtulec, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague, Czech Republic. Tel.: +420 296 442 256, Fax: +420 296 442 154; e-mail: trachtul@biomed.cas.cz.

Abbreviations: Chr. – chromosome, kb – kilobase pairs, Mb – megabase pairs, ORF – open reading frame.

Little is known about the extension of synteny or linkage conserved across large genetic distances, for example among human, *C. elegans*, Drosophila, *Anopheles gambiae*, *Arabidopsis thaliana*, yeast, and other eukaryotic genomes (Trachtulec and Forejt, 2000, for a review). Conserved linkage is defined as the presence of at least two pairs of homologous genes on the same chromosome in two species in the same order and orientation; conserved synteny is the presence of at least two pairs of homologous genes on the same chromosome in two species, regardless of their order (Nadeau and Sankoff, 1998). The term "conserved synteny" can therefore also be applied to genes that are not neighbours. The orthologous relationships among the homologues from the model organisms also remain largely unknown, despite the fact that most or all of their genome has already been sequenced (Goffeau et al., 1996; The *C. elegans* Sequencing Consortium, 1998; Adams et al., 2000; Arabidopsis Genome Initiative, 2000; Douglas et al., 2001; Katinka et al., 2001; Lander et al., 2001; Venter et al., 2001; Gardner et al., 2002; Glockner et al., 2002; Holt et al., 2002; Wood et al., 2002).

To gain an insight into the animal operons and into the genome organization of eukaryotes in general, the organization of homologues of operon members in other eukaryotic genomes was investigated.

Material and Methods

The names of the operon genes were taken from the literature and sequences extracted from the NCBI site. In order to define the closest homologues (orthologues), reciprocal sequence comparisons of the proteins encoded by operons with proteins from the nonredundant database were performed by the gapped BLASTP program (Altschul et al., 1997). If there was no better hit in the reference species than with a P value below 10^{-4}, the operon gene was viewed as nonconserved in this species. If the hits in all the reference species were of a lower value than to a protein from the same operon, the operon members were considered duplications. The resulting best hits for the reference species were extract-

ed and used in new separately conducted BLASTP searches. All hits from the same reference species as the query with a better score than for the operon product initially used were considered orthologous and inspected for identity. The names of the genes and chromosomal positions were then extracted from the Ensembl database (at www.ensembl.org) for the human hits and FlyBase (flybase.bio.indiana.edu) for the Drosophila hits. The chromosomal positions of orthologues of the genes from the same operon were examined for linkage and the positives evaluated statistically.

A simple formula published previously (Trachtulec and Forejt, 2001; Durand and Sankoff, 2003) was used to count the probability of the syntenies: $P = (r/G)^{(n-1)}$, where "r" is the region encompassed by the genes, "G" the genome size, and "n" is the number of the syntenic genes. The formula can be used only for very large genetic distances and only for genes or open reading frames tightly linked in at least one of the compared species. For multiple linked orthologues in the reference species, only the position of the orthologue most distant to the syntenic member was considered.

Results

Synteny analysis of C. elegans operon members

In total, 158 genes in 60 randomly chosen *C. elegans* operons were analysed for orthologues in two animal species, *Homo sapiens* and *Drosophila melanogaster*, by reciprocal BLAST searches (see Supplemental Table). Six operons (10%) containing 14 genes were the results of duplications. Another 13 operons (22%) contained 16 genes not conserved in the genomes of Drosophila nor human. The synteny thus could not be investigated for 12 operons (25%) composed solely of duplicated or nonconserved genes. Four of the 16 genes not conserved in fly or man resided in four operons, for which there were at least two other conserved members. Other six genes in five operons were conserved in man only; the human orthologues of the genes from these operons were not syntenic. There remained 41 operons (68%) encompassing 114 genes or 99 pairs of genes, which could be analysed for synteny both in Drosophila and in man. For 18 gene pairs (20% of 99) in 15 operons, there was a synteny in one reference species at a probability below 0.05 and, moreover, two gene pairs had syntenies in both species with $P < 0.01$ each. The first of the two pairs, genes for fibrillarin and ribosomal protein S16, was mapped to a 400-kilobase pairs (kb) region in human ($P = 0.00013$) and 466 kb in Drosophila (0.004). The second pair, encoding a proteasome subunit and a reductase gene, has homologues syntenic in a human region of 33.2 megabase pairs (Mb) in length ($P = 0.01$) and in a DNA-stretch of 240 kb in the fruit fly ($P = 0.002$). The number of the pairs expected to be found in the distance corresponding to less than 1% of the genome size in both human and Drosophila from 99 gene pairs, supposing random gene order, is 0.01 ($P_{observed/expected} < 0.0001$, binomial equation).

To increase the probability of finding significant conserved syntenies, all members of the 15 operons encompassing gene pairs syntenic in Drosophila or human were then compared by reciprocal BLAST searches to other eukaryotic genomes and their map positions were determined. The two pairs of operon genes with closely linked members in both Drosophila and human were found to be syntenic in other genomes (see below), increasing the probability that these pairs are under selection. From 12 pairs syntenic in Drosophila but not in human, only two pairs were also syntenic in *A. gambiae* with $P < 0.05$. From six pairs syntenic in human but not in Drosophila, four pairs were also syntenic in the mouse.

The most significant conserved syntenies of C. elegans operon members

The most significant synteny involves the genes for fibrillarin and ribosomal protein S16. The *C. elegans* genes *fib-1* and *rps-16* are in an operon on on chromosome (Chr.) V (Zorio et al., 1994, operon No. 15 in the Supplemental Table). The products of these genes participate in the function of the ribosome. Fibrillarin is a nucleolar protein involved in the modification of ribosomal RNA and the S16 protein is a part of the small ribosomal subunit. Its orthologue from *E. coli*, ribosomal protein S9, binds to ribosomal RNA (Wiener et al., 1988). The human fibrillarin gene (*FBL*) and ribosomal protein S16 gene (*RPS16*) were mapped to Chr. 19q13.1. The genes are single copy (excluding intronless pseudogenes), reside 400 kb apart and the probability of this arrangement is 0.0001. The mouse *Fbl* and *Rps16* genes map to a 177-kb region on mouse Chr. 7 ($P = 0.00006$). Drosophila *Fib* and CG4046 genes reside on Chr. 2R in a distance of 466 kb (0.004). The orthologues of these two genes also map on Chr. 3 of *A. gambiae* (5.3 Mb apart, $P = 0.02$), *S. pombe* Chr. IIR (*rps16* and *fib*, 1.2 Mb, 0.1), *G. theta* nucleomorph Chr. 3 (*rps16* and *nop1*, 32 kb, 0.06), and *P. falciparum* 3D7 Chr. 14 (PF14_0132 and PF14_0068, 274 kb, 0.01). The genome of *S. cerevisiae* encompasses two genes for ribosomal protein S16, one of which (*RPS16B*) is contained in Chr. IV, 120 kb apart from the single copy gene for fibrillarin (*NOP1*, $P = 0.02$). These genes are not linked in the genomes of *Candida albicans*, *Neurospora crassa*, *A. thaliana, Leishmania major* or *Encephalitozoon cuniculi*. None of the archeal genomes investigated carried the homologues of these two genes in the same operon.

The next significant gene pair, the short-chain acyl reductase (C15H11.4) and the proteasome subunit (alpha type 6, *pas-1*, C15H11.7), is encompassed in an operon of four genes (No. 7, Table 1). The two remaining genes are a small subunit of RNA polymerase I (C15H11.8) and a regulator of ribosome biosynthesis

Table 1. Conserved synteny analysis of operon No. 7

Function /Species	Ribosomal regulation	RNA polymerase I small subunit	Proteasome subunit	Reductase (dehydrogenase)
Cel	C15H11.9	C15H11.8	C15H11.7	C15H11.4
Hsa	KIAA0112 8q13.1	ZNRD1=RPA12 6p21.3	**PSMA6 14q13-33.1**	**FLJ30273,ARSDR1 14q24-66.2,66.3 Mb**
Mdo	1-9.6 Mb	17-36.1	**Psma6 12-49.9**	**Mtd1 12-73.8**
Dme	LD24586p 64F	RpI12 93F	**Prosalpha6 43E**	**5x 43E**
Aga	**Chr2-124.8 Mb**	**2-26.7**	**2-62.8**	**~8x Chr2-100 to 135**
Sce	**RRS1 XV**	RPA12 X	PRC2 VII	**YOR246C XV**
Spo	**SPBC29A3.16 II-2.03**	**rpa12 III-1.04**	**SPBC646.16 II-0.88**	**SPCC736.13 III-0.34**
Ath	At2g37990 2-15.9	MPE11.11, T20H2.16 3, 1	prc1, PAA2 5, 2-2.2	7x 1,2-15.7,4,5
Ecu	**ECU08_1510 VIII-0.17Mb**	**ECU08_1330 VIII-0.15**	ECU10_0550 X	nc
Ddi	**2-2.71 Mb**	2-?	2-?	**2-1.76 Mb**

Note: The name of the gene, if assigned, is followed by its chromosomal position; the significant syntenies are in **bold**; nc, not conserved; ?, not known; Cel, *C. elegans*; Hsa, *Homo sapiens*; Mdo, *Mus domesticus*; Dme, *Drosophila*; Aga, *Anopheles gambiae*; Ath, *Arabidopsis thaliana*; Spo, *S. pombe*; Sce, *S. cerevisiae*; Ecu, *Encephalitozoon cuniculi*; Ddi, *Dictyostelium discoideum*.

C15H11.9. The reductase genes are multicopy in human (2 copies in 100 kb), Drosophila (5 genes in 12 kb), *A. gambiae* (about 8 copies in 35 Mb), and *A. thaliana* (7 genes on four chromosomes). The orthologues of all four operon genes map to Chr. 2 of *A. gambiae* (P = 0.04) and probably all four to Chr. 2 of *Dictyostelium discoideum* (only one distance is known: P = 0.03). The proteasome subunit and reductase genes are syntenic on human Chr. 14 (P = 0.01), mouse Chr. 12 (0.008), and fruit fly Chr. 2 (0.002). The ribosomal regulator gene is syntenic with the proteasome subunit on *S. pombe* Chr. II (0.09), with the reductase gene on *S. cerevisiae* Chr. XV (0.006), and with the RNA polymerase subunit gene on *E. cuniculi* Chr. VIII (P = 0.008). The C15H11.8 and C15H11.4 homologues reside on *S. pombe* Chr. III (0.06). The C15H11.9 homologue maps with the C15H11.4 and *pas-1* genes to Chr. 2 of *A. thaliana* (P = 0.1).

The operon No. 29 (Table 2) encompasses genes for a proton-transporting ATPase ZK637.8, a thiamine pyrophosphokinase ZK637.9, and a thiredoxin reductase ZK637.10. Although four homologues of ZK637.8 and two homologues of ZK637.10 exist in human and mouse, there is synteny of *ATP6V0A2* and *TXNRD1* on human Chr. 12 (P = 0.03), and synteny of *ATP6N1B* and *TPK1* on Chr. 7 (0.007). All three operon genes have homologues on mouse Chr. 6 (0.001) and on Drosophila Chr. 3 (P = 0.14). The thioredoxin reductase is not conserved in yeast or plants.

Another operon (No. 8, Table 3) is composed of three genes encoding tubulin gamma (*tbg-1*), a 16-kDa subunit of RNA polymerases I and III (F58A4.9), and an ubiquitin-conjugating enzyme (F58A4.10). The homologues of *tbg-1* and F58A4.10 are linked on human Chr. 17 (P = 0.01), mouse Chr. 11 (0.01), *S. pombe* Chr. IIR (0.06), *A. thaliana* Chr. 3 (0.08), *D. discoideum* Chr. 2 (0.002), but not in *D. melanogaster*, *A. gambiae*, *N. crassa* or *E. cuniculi*.

Synteny analysis of operon members from other animals

Multicistronic transcription also occurs in human or Drosophila, but only one gene, human *MOCS1* for molybdenum cofactor biosynthesis, was analysed in

Table 2. Conserved synteny analysis of operon No. 29

Function /Species	Vacuolar H+ pump	Thiamine pyrophosphokinase	Thioredoxin reductase
Cel	ZK637.8	ZK637.9	ZK637.10
Hsa	ATP6N1A, **ATP6N1B**, ATP6V0A2, TCIRG1; 17q21, **7q34**, **12q24**, 11q13	**TPK1 7q35**	TXNRD1,2; **12q23**,22q11
Mdo	**Atp6n1a**, Atp6v0a1,2, Tcirg1; **6-38.2**, 11, 5, 19	**Tpk1 6-43.6 Mb**	**Txnrd1**,2; **6-90.5**, 10; 16
Dme	Vha100-1, CG12602, Vha100-2, CG7678; 99A, 32F; 91A	CG14721 86F	Trxr-1, Trxr-2; 7D, 79E
Aga	2-55.9,8.8;un	2-36.2	X-5.4
Sce	**VPH1**, STV1; **XV-0.8**, XIII	**THI80 XV-0.6**	nc
Spo	SPAC16E8.07c I	tnr3 I	nc
Ath	At2g21410, At4g39080; 2-9.1, 4	At2g44750, At1g02880; 2-18.4, 1	nc

Note: One gene pair has a significant synteny (each synteny P < 0.05); un, unassigned; see Table 1 for additional legend.

Table 3. Conserved synteny analysis of operon No. 8

Function /Species	Tubulin gamma	RNA polymerase I and III subunit, 16 kDa	Ubiquitin-conjugating enzyme
Cel	tbg-1(F58A4.8)	F58A4.9	F58A4.10
Hsa	**TUBG1,2; 17q21**	RPA16 13q12.2	**UBE2G1 17p13.3**
Mdo	**Tubg 11-101.9**	5-146	**Ube2g1 11-73.3**
Dme	**GammaTub 23C,37C**	**CG10685 37C**	CG9602,RE63412p 87F,un
Aga	3-1.5 Mb	3-50.5 Mb	2-23.3 Mb
Sce	TUB4 XIIR	YOL005c XV	?
Spo	SPBC32F12.04 IIR	rpb11 IL	SPBC1105.09 IIR
Ath	TUBG1,2; 3-22.8,5	At2g29540 2	UBC7,13 3-20.5, -17.1
Ecu	ECU08_0670 VIII	ECU11_0590 XI	ECU04_0630 IV
Ncr	Ctg212 I,VI	Ctg13 III,VI	Ctg421 IV
Cal	Ctg6-1674 un	Ctg6-2179 Chr. R	Ctg6-2511 Chr. R
Ddi	**dd_01918 2-3.53 Mb**	?	**dd_01380 2-3.58**
Gth	tubg 1	rpc9 3	?

Note: At least one gene pair has a significant synteny; Ctg, contig; Gth, *Guillardia theta* nucleomorph genome; Ncr, *Neurospora crassa*; Cal, *Candida albicans*; see Table 1 for additional legend.

distantly related animal species (Gray and Nicholls, 2000); its organization is conserved among vertebrates, as well as in Drosophila and *C. elegans*, but there are two nonsyntenic genes in *A. thaliana*.

The mapping positions of homologues of the multicistronic genes were obtained for other eukaryotes by reciprocal similarity searches. One of two open reading frames (ORFs) identified in a Drosophila bicistronic mRNA (Schulz et al., 1990) is a nonconserved protein. The same applies to other two bicistronic mRNAs of the fruit fly (Andrews et al., 1996; Ibnsouda et al., 1998). Both operon products can be nonconserved (Liu et al., 2000). The two ORFs carried on a single Drosophila mRNA (Brogna and Ashburner, 1997) are the result of a recent duplication. Both products of the human GDF1-LASS1 transcript (Lee, 1991) are conserved in nonvertebrate animals, but the homologues are not syntenic in Drosophila, *A. gambiae* or *C. elegans*. The SNURF product of the imprinted mammalian SNURF-SNRPN mRNA (Gray et al., 1999) is not conserved in nonvertebrates.

Another human gene for molybdenum cofactor biosynthesis, *MOCS2*, also produces a bicistronic mRNA (Reiss et al., 1999). Transcript sequences encoding both ORFs were found by TBLASTN searches in *Silurana tropicalis* (e.g., GenBank access. No. AL852203), *Oryzias latipes* (BJ007717), *Ciona intestinalis* (BW217776), *D. melanogaster* (AY047520, the gene is on Chr. 3R), and *A. gambiae* (BM625077, Chr. 2). However, the homologous ORFs map to two separate chromosomes in *C. elegans*, *D. discoideum*, *Emericella nidulans*, *A. thaliana*, or *Oryza sativa*. Thus, there are at least two genes producing bicistronic transcripts with their gene organization shared among eukaryotes, but the operon-like arrangement is not the only possible one.

Discussion

The data presented in this paper confirm the existence of syntenies of sequence-unrelated genes, which are conserved over large genetic distances. Unlike conserved syntenies known from among more closely related genomes (Nadeau and Sankoff, 1998, for a review), there are almost no syntenic segments (stretches of one conserved synteny not interrupted by any other conserved synteny). Therefore, a mathematical approach has to be employed to exclude a background. Two classical examples of nonrandom distribution of genes conserved in vertebrates and nonvertebrates are clusters of members related by homology, the homeobox and histone gene families (Maxson et al., 1983; Graham et al., 1989). The Drosophila Hox (HOM-C) homeobox cluster is split into two regions of Chr. 3R. Their distance, 10 Mb, represents about 8% of the Drosophila genome size (Adams et al., 2000). The *C. elegans* Hox genes are spread over 260 kb (0.3% of the genome), with many non-Hox genes interspersed among them (The C. elegans Sequencing Consortium, 1998). The *evenskipped* homeobox gene homologues are tightly linked to the human and cnidarian Hox genes (Pollard and Holland, 2000; Popovici et al., 2001) mapped 3.1 Mb to the *C. elegans* Hox cluster ($P = 0.03$), but are located on a different chromosome than the fruit fly or mosquito Hox genes. These data suggest that a nonrandom gene distribution or conserved synteny should be present in at least three distantly related eukaryotic genomes with the probabilities below 0.05. The probability can be calculated only for genes very tightly linked in one of the

species (e.g., operon members) with the formula presently available. A more sophisticated mathematical approach may be necessary to construct syntenic maps of distantly related model organisms. It should also be noted that synteny was not investigated for non-operonic *C. elegans* genes and the frequency of their syntenies cannot thus be compared to the frequency of syntenies of operonic genes.

At least five pairs of genes (5% of 99 investigated) of four operons (10% of 41) in our analysis would pass the rules defined above. The nonrandom arrangement of orthologues was also found for only two of the nine other animal operons investigated. This fact suggests that an extensive random shuffling takes place in shaping the eukaryotic genomes and only a small part can be under selection. Syntenies of genes of unrelated sequence conserved in mammals and nonvertebrates (Trachtulec et al., 1997; Pebusque et al., 1998; Abi-Rached et al., 2002) and even in the fission yeast *S. pombe* (Trachtulec and Forejt, 2001) have been found previously. The presence of the conserved synteny of the *PDCD2*, *TBP*, and *PSMB1* genes has also been recently confirmed in nonmammalian vertebrates (Trachtulec, Vlček, Mihola, Forejt, submitted). Cases of no synteny of gene homologues on larger genetic distances have also been published (reviewed in Trachtulec and Forejt, 2000, 2001). The examples have been too few to say how much of the synteny is conserved across eukaryotes. Taken together with the data presented here, however, it seems that significantly conserved syntenies are rare among distantly related eukaryotic genomes.

The fact that only syntenies but almost no gene order or syntenic segments remain conserved implies that intrachromosomal rearrangements occur more often than interchromosomal events. This view is supported by the data from syntenies shared among more closely related genomes (Nadeau and Sankoff, 1998; McLysaght et al., 2000). We can only speculate about the reason behind it. Nevertheless, there are at least two possible explanations. First, the semisterility of translocation heterozygotes is a much more serious defect of the reproductive fitness of an individual than the suppression of recombination, the usual consequence of inversions. Second, there can be shared regulatory elements among the syntenic genes, which would not be accessible after a translocation, for example due to spatial constraints in the nucleus of the cell. The distances of these elements from the regulated genes can be more than 800 kb (reviewed by Kleinjan and van Heyningen, 1998). This view is also supported by the discovery of chromosomal territories (Zink et al., 1998).

Acknowledgements

The author wishes to thank Dr. J. Forejt, Dr. J. Pačes and P. Strnad for helpful comments.

References

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P., Inoko, H. (2002) Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.* **31,** 100-105.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000) The genome sequence of Drosophila melanogaster. *Science* **287,** 2185-2195.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389-3402.

Andrews, J., Smith, M., Merakovsky, J., Coulson, M., Hannan, F., Kelly, L. E. (1996) The stoned locus of Drosophila melanogaster produces a dicistronic transcript and encodes two distinct polypeptides. *Genetics* **143,** 1699-1711.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408,** 796-815.

Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M., Kim, S. K. (2002) A global analysis of Caenorhabditis elegans operons. *Nature* **417,** 851-854.

Brogna, S., Ashburner, M. (1997) The Adh-related gene of Drosophila melanogaster is expressed as a functional dicistronic messenger RNA: multigenic transcription in higher organisms. *EMBO J.* **16,** 2023-2031.

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L. T., Wu, X., Reith, M., Cavalier-Smith, T., Maier, U. G. (2001) The highly reduced genome of an enslaved algal nucleus. *Nature* **410,** 1091-1096.

Durand, D., Sankoff, D. (2003) Tests for gene clustering. *J. Comput. Biol.* **10,** 453-482.

Evans, D., Zorio, D., MacMorris, M., Winter, C. E., Lea, K., Blumenthal, T. (1997) Operons and SL2 trans-splicing exist in nematodes outside the genus Caenorhabditis. *Proc. Natl. Acad. Sci. USA* **94,** 9751-9756.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S. et al. (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419,** 498-511.

Glockner, G., Eichinger, L., Szafranski, K., Pachebat, J. A., Bankier, A. T., Dear, P. H., Lehmann, R., Baumgart, C., Parra, G., Abril, J. F. et al. (2002) Sequence and analysis of chromosome 2 of Dictyostelium discoideum. *Nature* **418,** 79-85.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M. et al. (1996) Life with 6000 genes. *Science* **274,** 546, 563-567.

Graham, A., Papalopulu, N., Krumlauf, R. (1989) The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell* **57,** 367-378.

Gray, T. A., Nicholls, R. D. (2000) Diverse splicing mechanisms fuse the evolutionarily conserved bicistronic MOCS1A and MOCS1B open reading frames. *RNA* **6,** 928-936.

Gray, T. A., Saitoh, S., Nicholls, R. D. (1999) An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc. Natl. Acad. Sci. USA* **96,** 5616-5621.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M., Wides, R. et al. (2002) The genome

sequence of the malaria mosquito Anopheles gambiae. *Science* **298,** 129-149.

Ibnsouda, S., Ferrer, P., Vincent, A. (1998) Conservation of read-through transcription of the Drosophila serendipity genes during evolution is gratuitous. *Mol. Gen. Genet.* **259,** 484-490.

Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P. et al. (2001) Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. *Nature* **414,** 450-453.

Kleinjan, D. J., van Heyningen, V. (1998) Position effect in human genetic disease. *Hum. Mol. Genet.* **7,** 1611-1618.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409,** 860-921.

Lee, S. J. (1991) Expression of growth/differentiation factor 1 in the nervous system: conservation of a bicistronic structure. *Proc. Natl. Acad. Sci. USA* **88,** 4250-4254.

Liu, H., Jang, J. K., Graham, J., Nycz, K., McKim, K. S. (2000) Two genes required for meiotic recombination in Drosophila are expressed from a dicistronic message. *Genetics* **154,** 1735-1746.

Maxson, R., Cohn, R., Kedes, L., Mohun, T. (1983) Expression and organization of histone genes. *Annu. Rev. Genet.* **17,** 239-277.

McLysaght, A., Seoighe, C., Wolfe, K. H. (2000) High frequency of inversions during eukaryote gene order evolution. In: *Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and the Evolution of Gene Families,* eds. Sankoff, D., Nadeau, J. H., pp. 47-58, Kluwer Academic Publisher, Dordrecht.

Nadeau, J. H., Sankoff, D. (1998) Counting on comparative maps. *Trends Genet.* **14,** 495-501.

Pebusque, M. J., Coulier, F., Birnbaum, D., Pontarotti, P. (1998) Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. *Mol. Biol. Evol.* **15,** 1145-1159.

Pollard, S. L., Holland, P. W. (2000) Evidence for 14 homeobox gene clusters in human genome ancestry. *Curr. Biol.* **10,** 1059-1062.

Popovici, C., Leveugle, M., Birnbaum, D., Coulier, F. (2001) Homeobox gene clusters and the human paralogy map. *FEBS Lett.* **491,** 237-242.

Reiss, J., Dorche, C., Stallmeyer, B., Mendel, R. R., Cohen, N., Zabot, M. T. (1999) Human molybdopterin synthase gene: genomic structure and mutations in molybdenum cofactor deficiency type B. *Am. J. Hum. Genet.* **64,** 706-711.

Schulz, R. A., Miksch, J. L., Xie, X. L., Cornish, J. A., Galewsky, S. (1990) Expression of the Drosophila gonadal gene: alternative promoters control the germ-line expression of monocistronic and bicistronic gene transcripts. *Development* **108,** 613-622.

The C. elegans Sequencing Consortium (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282,** 2012-2018.

Trachtulec, Z., Forejt, J. (2000) Syntenies of unrelated genes conserved in mammals and nonvertebrates (a review). In: *Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and the Evolution of Gene Families,* eds. Sankoff, D., Nadeau, J. H., pp. 357-366. Kluwer Academic Publishers, Dordrecht.

Trachtulec, Z., Forejt, J. (2001) Synteny of orthologous genes conserved in mammals, snake, fly, nematode, and fission yeast. *Mamm. Genome* **12,** 227-231.

Trachtulec, Z., Hamvas, R. M., Forejt, J., Lehrach, H. R., Vincek, V., Klein, J. (1997) Linkage of TATA-binding protein and proteasome subunit C5 genes in mice and humans reveals synteny conserved between mammals and invertebrates. *Genomics* **44,** 1-7.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A. et al. (2001) The sequence of the human genome. *Science* **291,** 1304-1351.

von Mering, C., Bork, P. (2002) Teamed up for transcription. *Nature* **417,** 797-798.

Wiener, L., Schuler, D., Brimacombe, R. (1988) Protein binding sites on Escherichia coli 16S ribosomal RNA; RNA regions that are protected by proteins S7, S9 and S19, and by proteins S8, S15 and S17. *Nucleic Acids Res.* **16,** 1233-1250.

Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S. et al. (2002) The genome sequence of Schizosaccharomyces pombe. *Nature* **415,** 871-880.

Zink, D., Cremer, T., Saffrich, R., Fischer, R., Trendelenburg, M. F., Ansorge, W., Stelzer, E. H. (1998) Structure and dynamics of human interphase chromosome territories in vivo. *Hum. Genet.* **102,** 241-251.

Zorio, D. A., Cheng, N. N., Blumenthal, T., Spieth, J. (1994) Operons as a common form of chromosomal organization in C. elegans. *Nature* **372,** 270-272.