Update TRENDS in Genetics Vol.21 No.1 January 2005

Research Focus

A bunch of fun-guys: the whole-genome view of yeast evolution

Howard Ochman[1], Vincent Daubin[1] and Emmanuelle Lerat[2]

[1]Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 87521, USA
[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 87521, USA

The resolution of the complete sequences of several hemiascomycete genomes provides new insights into the ways that yeast genomes change in size and in gene contents. These genomes provide evidence of whole-genome duplication occurring before the divergence of *Saccharomyces cerevisiae* and *Candida glabrata*, followed by massive gene loss that restored diploidy. The pattern of genome evolution in yeast differs from that in bacteria apparently as a result of stronger selective constraints on bacterial chromosomes.

The view that bacterial genomes evolved by successive doublings was appealing [1,2]. It seemed reasonable to consider small, simple genomes, such as those of the mycoplasmas, as primitive and that the larger, more complex forms arose subsequently through whole-genome duplications. This scenario explained the more-or-less discrete size classes of the bacterial genomes that were first available and also the orthogonal positions of some interrelated (and possibly duplicated) genes in certain bacteria with larger genomes [3,4]. But once the relationships of bacteria were evaluated from a molecular-phylogenetic perspective, all notions of whole-genome duplications in bacteria were dispelled: every clade of small-genome bacteria – those with genome sizes <1 Mb – is derived from relatives with larger genomes. No examples of whole-genome amplification have been found in the numerous completed bacterial-genome sequences; instead, it appears that the forces responsible for introducing new genes and increasing bacterial-genome size involve small segmental duplications, and more notably, lateral gene transfer.

By contrast, there is persuasive evidence that whole-genome duplication (WGD) had a significant role in the evolution of eukaryotic lineages including yeasts. Based on the arrangement, orientation and patterns of divergence of 376 gene pairs within 55 duplicated regions of the *Saccharomyces cerevisiae* genome, Wolfe and Shields [5] proposed that this yeast was an ancient polyploid, formed by a WGD event that occurred subsequent to the split from *Kluyveromyces lactis* some 100–200 million years ago. Surprisingly, the gene pairs remaining after this duplication constitute only 12% of the coding capacity of *S. cerevisiae*, indicating that most paralogous copies were subsequently lost during the evolution of the contemporary diploid genome.

Whole-genome duplication

The recent determination of the full genome sequences of several other yeasts has served to pinpoint the time of this WGD and, most notably, revealed the manner in which the ancient tetraploid genome degraded in the lineage leading to *Saccharomyces*. When the gene order contents of *S. cerevisiae* are compared with those of either *Ashbya gossypii* [6] or *Kluyveromyces waltii* [7], both of which diverged before the proposed WGD event (Figure 1), a curious pattern emerges: genes within a single block in each of these genomes were found to alternate homologs with two dispersed syntenic regions in the *S. cerevisiae* genome. These syntenic regions have two matches in *S. cerevisiae*, the expected result of WGD; but what is remarkable is that the two homologous regions in the *S. cerevisiae* genome rarely have any genes in common. This implies that after duplication of the entire genome, one copy of each paralogous gene pair was eliminated, apparently at random, by several small deletions averaging approximately two genes in length, thereby largely restoring the original ploidy and gene content. Interestingly, the duplicated genes that are present in the *Saccharomyces* genome are biased towards those with low evolutionary rates in other lineages [8]. It is notable that among these duplicates, there are several examples where one of the paralogs has assumed a new function [7], suggesting that new and useful traits are most easily derived from duplicated copies of conserved and/or essential genes.

Gene duplication has long been recognized as a key mechanism in the evolution of new functions [9]; but the process that occurred in the *S. cerevisiae* lineage, whereby WGD is followed by the massive elimination of individual genes, seems an unusual means for generating genetic novelty because so few duplicates are retained. The recent publication of four new hemiascomycetes genomes has revealed that other yeasts have reached comparable or higher degrees of genome redundancy via different mechanisms. By comparing these genomes in a phylogenetic context, Dujon *et al.* [10] proposed a scenario for the evolution of hemiascomycete genomes in which most new genes are generated by tandem, block or segmental duplications, rather than by WGD. Moreover, their analysis also reveals the fate of duplicated genes in the various yeast genomes. For example, *Candida glabrata*,

Corresponding author: Ochman, H. (hochman@email.arizona.edu).
Available online 18 November 2004

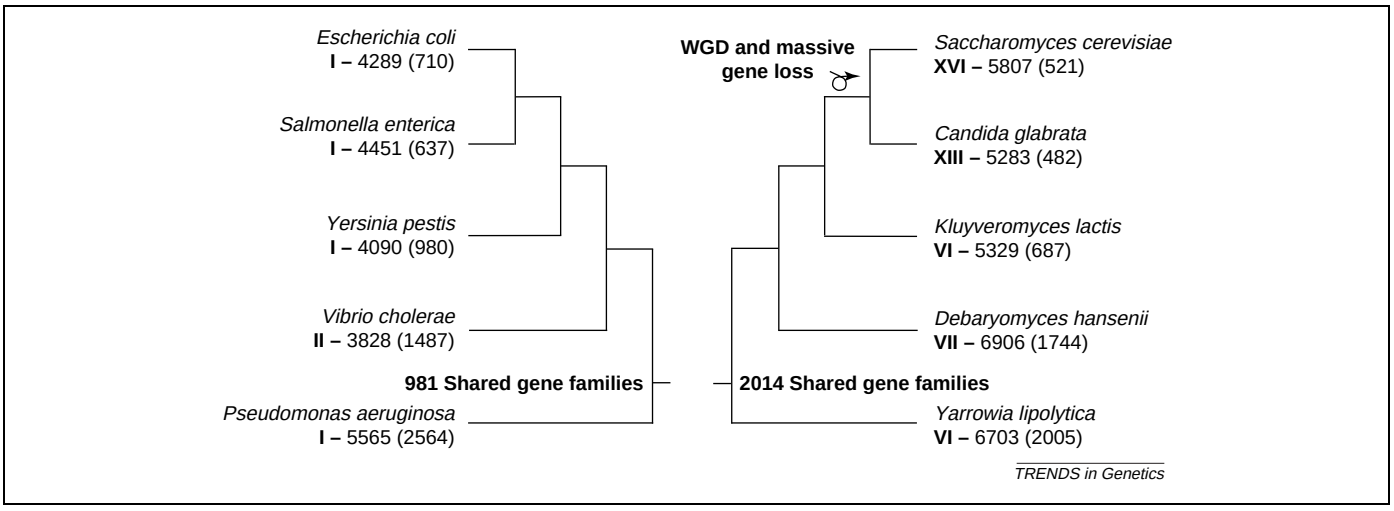

Figure 1. A comparison of genome evolution in bacteria and yeast. Phylogenetic relationships (and the numbers or criteria for defining gene families) are based on Lerat *et al.* [20] for gamma-proteobacteria and on Dujon *et al.* [10] for hemiascomycete yeasts. Under each species name, its chromosome number is indicated in roman numerals, followed by the total number of genes in each genome, with the number of species-specific genes given in parentheses. The timing of whole-genome duplication (WGD) in the lineage leading to *Saccharomyces cerevisiae* is shown, and the numbers of gene families shared among either the bacterial or the yeast genomes considered are indicated at the base of the tree.

which split from the *Saccharomyces* lineage after the WGD event, displays an even lower degree of genome redundancy than *S. cerevisiae*, possibly as a result of ongoing genome reduction related to its parasitic lifestyle.

Genome evolution in yeast

The emerging picture of genome evolution in hemiascomycetes, which, in the words of Dujon *et al.* [10], 'encompass an evolutionary span as large as the entire phylum of chordates' turns out to be distinct from that of bacteria, the only other group for which comparable genome information is currently available. To illustrate these differences, we selected five gamma-proteobacterial genomes (Figure 1), whose phylogenetic relationships and overall levels of protein divergence mirror those of the yeast genomes considered by Dujon *et al.* [10]. Despite the fact that bacteria generally contain fewer genes than yeast and might be thought to share numerous essential genes, only 10% (981 of 9853, representing 33% of the total number of genes) of the protein families detected in these bacteria are common to all species, whereas >20% (2014 of 8983, representing 57% of the total number of genes) of the yeast protein families were present in all genomes examined. However, the proportion of proteins restricted to a single species is similar in both groups and often accounts for a large fraction of a genome, reaching 30% in *Yarrowia lipolytica* and >40% in *Pseudomonas aeruginosa*. The source of such genes has been examined in bacteria and is best explained by gene acquisition from phylogenetically distant sources [11]. But whether the majority of unique genes in these yeast genomes arose by lateral gene transfer or by the divergence of native sequences is unresolved.

Despite the relatively high degree of overlap in the gene inventories of the sequenced hemiascomycetes, their genomes display more rearrangements than those detected in related bacteria. The two most closely related yeast, *S. cerevisiae* and *C. glabrata*, share several thousand genes, which are dispersed into >500 syntenic clusters. By contrast, the gene order in the chromosomes of the bacteria *Escherichia coli* and *Salmonella enterica* – whose split, based on the extent nucleotide-sequence divergence, most likely occurred earlier than that of *S. cerevisiae* and *C. glabrata* – differ by only a single inversion.

Concluding remarks

Although chromosome number and geometry can vary among bacteria, their genome organization is under much stronger selection than yeast genomes. Several factors appear to contribute to the conservation of gene position and order in bacteria including the arrangement of genes into operons [12], gene dosage and orientation [13–15], the relative positions of the replication origin and terminus [16–19] and the reliance on a single replication origin per chromosome. Variation in these constraints, coupled with the diverse selective pressures that are induced by the reproductive mode, population structure and lifestyle of an organism are expected to generate high levels of genomic diversity. Although currently limited to unicellular organisms with small, gene-rich genomes, the analysis of complete genome sequences in a phylogenetic context, such as that performed by Dujon *et al.* [10], will ultimately divulge the specific mechanisms contributing to the contents and organization of genomes.

References

1 Wallace, D.C. and Morowitz, H.J. (1973) Genome size and evolution. *Chromosoma* 40, 121–126
2 Sparrow, A.H. and Nauman, A.F. (1976) Evolution of genome size by DNA doublings. *Science* 192, 524–529
3 Hopwood, D.A. (1967) Genetic analysis and genome structure in *Streptomyces coelicolor*. *Bacteriol. Rev.* 31, 373–403
4 Riley, M. and Anilionis, A. (1978) Evolution of the bacterial genome. *Annu. Rev. Microbiol.* 32, 519–560
5 Wolfe, K.H. and Shields, D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387, 708–713
6 Dietrich, F.S. *et al.* (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 304, 304–307

7 Kellis, M. *et al*. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428, 617–624
8 Davis, J.C. and Petrov, D.A. (2004) Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol.* 2, E55
9 Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag
10 Dujon, B. *et al*. (2004) Genome evolution in yeasts. *Nature* 430, 35–44
11 Daubin, V. and Ochman, H. (2004) Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* 14, 1036–1042
12 Huynen, M.A. *et al*. (2000) Exploitation of gene context. *Curr. Opin. Struct. Biol.* 10, 366–370
13 Schmid, M.B. and Roth, J.R. (1987) Gene location affects expression level in *Salmonella typhimurium*. *J. Bacteriol.* 169, 2872–2875
14 Segall, A. *et al*. (1988) Rearrangement of the bacterial chromosome: forbidden inversions. *Science* 241, 1314–1318
15 Rocha, E.P. (2004) The replication-related organization of bacterial genomes. *Microbiology* 150, 1609–1627
16 Liu, S.L. and Sanderson, K.E. (1995) Rearrangements in the genome of the bacterium *Salmonella typhi*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 1018–1022
17 Eisen, J.A. *et al*. (2000) Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* 1, RESEARCH0011 doi:10.1186/gb-2000-1-6-research0011 (http://genomebiology.com/2000/1/6/research/0011)
18 Suyama, M. and Bork, P. (2001) Evolution of prokaryotic gene order: genome rearrangements in closely related species. *Trends Genet.* 17, 10–13
19 Tillier, E.R. and Collins, R.A. (2000) Genome rearrangement by replication-directed translocation. *Nat. Genet.* 26, 195–197
20 Lerat, E. *et al*. (2003) From gene trees to organismal phylogeny in prokaryotes: The case of the gamma-Proteobacteria. *PLoS Biol.* 1, 101–119

0168-9525/$ - see front matter
doi:10.1016/j.tig.2004.11.006

X for intersection: retrotransposition both on and off the X chromosome is more frequent

Pavel P. Khil, Brian Oliver and R. Daniel Camerini-Otero

Genetics and Biochemistry Branch and Laboratory of Cellular and Developmental Biology,
National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

As the heteromorphic sex chromosomes evolved from a pair of autosomes, the sex chromosomes became increasingly different in gene content and structure from each other and from the autosomes. Although recently there has been progress in documenting and understanding these differences, the molecular mechanisms that have fashioned some of these changes remain unclear. A new study addresses the differential distribution of retroposed genes in human and mouse genomes. Surprisingly, chromosome X is a major source and a preferred target for retrotransposition.

Sex determination systems are different between species [1,2]. Many organisms have heteromorphic sex chromosomes that have a central role in sex determination. Because sex chromosomes are relatively recent evolutionary acquisitions they are particularly plastic [2]. Recent data indicate that, in addition to sex determination, many important features distinguish sex chromosomes from autosomes. For the Y chromosome, the structural divergence from the autosomes is obvious – >30% of the Y heterochromatin is composed of duplications and other rearrangements that are, for the most part, nearly identical in sequence [3]. These identical sequences are believed to be essential for maintaining the integrity of the coding sequences by intrachromosomal gene conversion. In addition, the Y chromosome contains a reduced and specific set of genes or is even absent in some species [3,4].

Corresponding author: Camerini-Otero, R.D. (camerini@ncifcrf.gov).
Available online 19 November 2004

Although chromosome X is similar to the autosomes in structure, some clear differences are evident [5,6]. For example, it appears that some functional classes of genes, such as those primarily expressed in brain and muscle, are enriched on the X chromosome [5,6]. One particularly interesting class of genes is the sex-biased genes. Evolutionary models and experimental data (Box 1; Figure 1) suggest that SEX-BIASED GENES (see Glossary) should be differentially represented on the sex chromosomes [7,8].

Retrotransposition and the X chromosome

Although the differential representation of sex-biased genes on the X chromosome is increasingly well documented, the molecular mechanisms underlying the gene movements responsible are less clear. The recently published study by Emerson and coauthors [9], describing the genomic history of retroposed genes in mammals, sheds some light on these gene movements to and from the

Glossary

Meiotic sex chromosome inactivation (MSCI): condensation of sex chromosomes at the onset of male meiosis. The mechanism of MSCI is different from that for random X-chromosome inactivation observed in females.
Sexual antagonism: when a gene is beneficial to one sex but harmful to the other.
Sex-biased genes: genes having differences in expression between sexes.
Retrogenes: functional genes generated by retrotransposition.
Retropseudogenes: retrotransposed genes that have lost their function.
LINE-1 (L1): an abundant family of non-long terminal repeat (LTR) retrotransposons that is present in the genomes of mammals.