

Functional divergence of gene duplicates through ectopic recombination

Joaquin F. Christiaens^{1,2*}, Sebastiaan E. Van Mulders^{3*}, Jorge Duitama^{1,2*}, Chris A. Brown^{1,2,4,5}, Maarten G. Ghequire¹, Luc De Meester⁶, Jan Michiels¹, Tom Wenseleers⁶, Karin Voordeckers^{1,2} & Kevin J. Verstrepen^{1,2+}

¹Department of Microbial and Molecular Systems, Centre of Microbial and Plant Genetics (CMPG), ²VIB Laboratory of Systems Biology, ³Centre for Malting and Brewing Science, Department of Microbial and Molecular Systems, Faculty of Bioscience Engineering, Leuven, Belgium, ⁴Faculty of Arts and Sciences Center for Systems Biology, Harvard University, Cambridge, ⁵Fathom Information Design, Boston, Massachusetts, USA, and ⁶Department of Biology, Animal Ecology and Systematics Section, Leuven, Belgium

Gene duplication stimulates evolutionary innovation as the resulting paralogs acquire mutations that lead to sub- or neofunctionalization. A comprehensive *in silico* analysis of paralogs in *Saccharomyces cerevisiae* reveals that duplicates of cell-surface and subtelomeric genes also undergo ectopic recombination, which leads to new chimaeric alleles. Mimicking such intergenic recombination events in the *FLO* (flocculation) family of cell-surface genes shows that chimaeric *FLO* alleles confer different adhesion phenotypes than the parental genes. Our results indicate that intergenic recombination between paralogs can generate a large set of new alleles, thereby providing the raw material for evolutionary adaptation and innovation.

Keywords: new genes; innovation; evolution; adhesins; flocculation

EMBO reports (2012) 13, 1145–1151. doi:10.1038/embor.2012.157

INTRODUCTION

How organisms evolve and adapt to new environments remains a central question in biology. Gene duplication events have a

crucial role in evolutionary processes, especially in the rapid development of new functions [1,2]. Duplication might yield adaptive benefits through increased dosage of the parental gene. In addition, gene duplication can also stimulate evolutionary innovation as mutations in one or both gene duplicates can lead to subfunctionalization or neofunctionalization [3–8].

Gene duplications are not spread evenly over the genome, occurring much more frequently in the subtelomeres, regions directly adjacent to the telomeres [9–11]. As a result, subtelomeric gene families are often large, with some families carrying as many as several hundred paralogs [9,11]. In *Trypanosoma* and *Plasmodium* species, variable expression of subtelomeric variants of a cell-surface antigen allows these pathogens to elude the host immune system. Subtelomeric gene families in the more compact genome of *Saccharomyces cerevisiae* are generally smaller and are enriched for cell-surface genes, as well as genes involved in nutrient transport and metabolism [9,12–18].

Recombination between paralogs could generate more sequence diversity in duplicated genes [19]. However, apart from anecdotal examples in specific gene families such as the *VAR* and *VSG* cell-surface genes of *Plasmodia* and *Trypanosomes* and the major histocompatibility complex (MHC) class genes in vertebrates, the occurrence and biological relevance of ectopic recombination between gene duplicates have not been systematically investigated [16,17,20].

Here, we present the results of a comprehensive *in silico* analysis of ectopic recombination in all paralog gene families in the model eukaryote *S. cerevisiae*. Our results show that intergenic recombinations occur predominantly in gene families that are located at the subtelomeres and/or encode cell-surface genes. To verify whether these intergenic recombination events could lead to altered phenotypes, we mimicked the intergenic recombination events that shaped the *FLO* (flocculation) adhesin gene family. Phenotypic analyses of these artificial *FLO* chimaera

¹Department of Microbial and Molecular Systems, Centre of Microbial and Plant Genetics (CMPG),

²VIB Laboratory of Systems Biology,

³Department of Microbial and Molecular Systems, Centre for Malting and Brewing Science, Faculty of Bioscience Engineering, KU Leuven, Kasteelpark Arenberg 22, B-3001 Leuven (Heverlee), Belgium

⁴Faculty of Arts and Sciences Center for Systems Biology, Harvard University,

52 Oxford Street, Cambridge, Massachusetts 02138, USA

⁵Fathom Information Design, 214 Cambridge Street, Boston, Massachusetts 02114, USA

⁶Department of Biology, Animal Ecology and Systematics Section, KU Leuven, Charles Deberiotstraat 32, B-3000 Leuven, Belgium

*These authors contributed equally to this paper.

+Corresponding author. Tel: +32 (0) 16 75 13 90; Fax: +32 (0) 16 75 13 91;

E-mail: kevin.verstrepen@biw.vib-kuleuven.be

Received 2 September 2012; accepted 17 September 2012; published online 16 October 2012

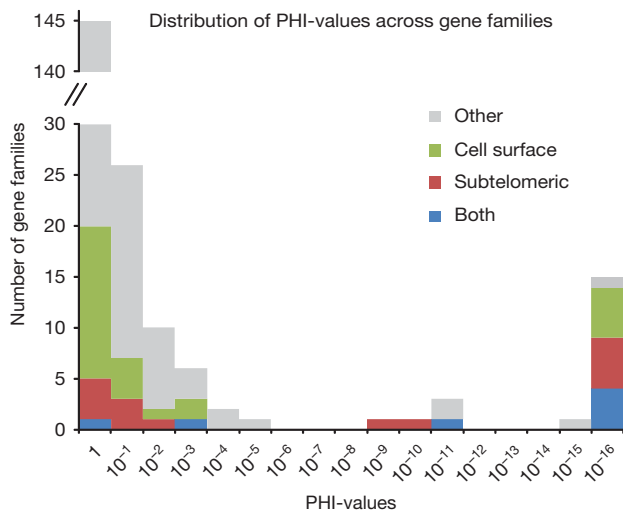


Fig 1 | Analyses of paralog gene families indicate intergenic recombination. Distribution of the PHI values for all paralog gene families in the *S. cerevisiae* genome. Most gene families cluster towards the left hand side of the graph, with high PHI values that are indicative of the absence of intergenic recombination. However, several gene families on the far right hand side (Table 1) show inter-paralog recombination. Note that this group exists almost exclusively of subtelomeric and/or cell-surface gene families. For more information on the calculation of PHI values, see Bruen *et al* [23] and (supplementary information online) methods. PHI, pairwise homoplasy index.

revealed that they were functional and conferred phenotypes that differed from their parental adhesins.

RESULTS AND DISCUSSION

Ectopic recombination in gene families

To investigate the occurrence of intergenic recombination events in *S. cerevisiae*, we first identified paralogous genes by BLASTing the reference strain S288c's proteome against itself. Next, we used the MCL clustering algorithm and manual curation to define a list of 210 gene families, each comprising at least two paralogs (see methods and supplementary information online for details) [21]. For all genes in these families we collected the known GO categories as well as chromosomal locations. We attributed these characteristics to gene families if at least two genes in a family shared the annotation (the rationale being that it takes at least two genes to create a chimaeric allele).

Next, we BLAST-searched each family against a database containing 24 recently published high-quality *S. cerevisiae* genomes (supplementary Table S1 online). For each family, we performed several *in silico* tests to check for evidence of intergenic recombination in the family. First, we used the SplitsTree4 programme [22] to produce reticulate phylogenetic trees of each family. In the absence of recombination, this procedure results in a classic, unrooted phylogenetic tree. Alternatively, when a recombination event took place, this is represented in the reticulate tree as a closed rectangle. To further analyse these families and provide statistically significant proof for the occurrence of intergenic recombination, we used the Pairwise Homoplasy Index (PHI) test [23]. The null hypothesis for this test is that the observed sequence differences are due to convergent mutations, which implies that all

PHI values are similar for all pairs (i.e., the PHI values do not vary with the physical distance between residues). Alternatively, in the presence of recombination, distant sites show low PHI values. These calculations then result in a single PHI value for the paralog gene family. For further details about these procedures, please refer the supplementary information online.

The distribution of PHI values indicates the presence of certain highly recombinogenic families (Fig 1). Interestingly, almost all chimaeric sequences are part of families that contain either subtelomeric or cell-surface genes, or both (Table 1). Statistical analyses of these distributions with a Komolgorov–Smirnov test revealed a statistically significant enrichment for subtelomeric gene families (P -value 3.48×10^{-6}) and for cell-surface gene families (P -value 5.5×10^{-3}). The distribution of a combination of all subtelomeric and cell-surface gene families was also significantly enriched for chimaeric alleles. Moreover, when all cell-surface gene families are removed from the analysis, we still find a statistically higher occurrence of chimaeric subtelomeric genes and *vice versa*. We were unable to find other subgroups that are enriched or depleted for ectopic recombination (e.g., the P -value for the comparison between families containing intragenic tandem repeats, and those without is 0.38; supplementary Fig S1 online).

Several recombinations shaped the adhesin gene family

To investigate the functional implications of ectopic recombination events, we focused on the subtelomeric *FLO* gene family, which encodes lectin-like cell-surface adhesins that confer adhesion to abiotic surfaces and/or other yeast cells [24–29]. These two phenotypes are biologically important for yeast cells and are relatively easy to measure and quantify [24,29–33]. The amino terminus of these rod-shaped proteins is a lectin-like globular domain that contains a pentapeptide involved in adhesion to specific carbohydrate residues present at the surface of other yeast cells or host tissues [34,35]. The central adhesin domain is formed by a repetitive pattern of a heavily glycosylated serine/threonine-rich peptide, which is thought to act as a variable rod-like spacer that helps to display the N-terminal domain to the environment [36].

To perform an in-depth analysis of recombination events between *FLO* paralogs, we gathered 58 more *FLO* sequences (including a few partial sequences or pseudogenes) from both NCBI and ENA (supplementary Table S2 online). Phylogenetic analysis revealed three subclades that cluster with the *FLO1*, *FLO10* and *FLO11* genes found in the reference S288c strain. In line with previous findings, we found extensive variation in tandem repeat length, even among adhesins from the same class [36,37].

We first verified whether the full set of *FLO* genes shows signs of intergenic recombination across the repeat region. Unrooted phylogenetic trees of the N- (Fig 2A) and carboxy-terminal (Fig 2B) domains revealed that several alleles might originate from recombination across the repeat region (Figs 2A–C). The PHI test further confirmed the occurrence of several intergenic recombination events in the complete *FLO* open reading frames (P -value $< 10^{-16}$). Whereas this shows that ectopic recombination between intragenic tandem repeats does occur, analyses using all *S. cerevisiae* families do not show a significant enrichment of chimaera among genes that contain internal tandem repeats (see above).

Table 1 | *S. cerevisiae* gene families showing ectopic recombination

Family description	Sequences in S288c	Total number of sequences retrieved	Subtelomeric	Cell-surface
Genes coding for (putative) Helicase-like proteins	31	323	Yes	No
<i>HXT</i> family	16	213	Yes	Yes
<i>COS</i> family	9	116	Yes	Yes
Genes coding for (Iso)maltases	7	86	Yes	No
<i>AAD</i> gene family	7	74	Yes	No
<i>FLO</i> gene family	6	19	Yes	Yes
Type I transmembrane sorting receptor for vacuolar hydrolases and similar sequences	6	69	Yes	No
<i>PHO</i> gene family	5	79	Yes	Yes
Pheromone-regulated protein with a motif involved in COPII binding/ putative integral membrane protein	4	27	No	Yes
Genes coding for MAL activators	4	75	Yes	No
<i>ENA</i> gene family	3	19	No	Yes
<i>TDH</i> gene family	3	52	No	No
<i>FRE</i> gene family	3	55	No	Yes
Transporters of thiamine/nicotinamide riboside	3	44	No	Yes
<i>TPO</i> gene family	2	24	No	Yes

FLO, flocculation. Examples of some representative *S. cerevisiae* gene families that undergo ectopic recombination (for a full list, see supplementary Dataset S1 online). The table contains 15 gene families that contain at least one chimaeric sequence (PHI value $< 10^{-16}$). First column lists function of characterized members of the identified gene family, 'sequences in S288c' lists the number of members of each gene family present in the S288c reference genome, whereas 'sequences retrieved' lists represents the total number of sequences found in 24 genomes (supplementary Table S1 online). Gene families are classified as subtelomeric or cell-surface if at least two sequences in the family meet the characteristic.

As repeat regions can cause artifacts in alignments, we repeated the analyses to specifically search for recombination events in the N-terminal domains of the *FLO* genes. These domains contain several sites important for the recognition and binding of specific carbohydrates [34,35], which is important for pathogenicity in *Candida* strains [38,39], and for *FLO* characteristics in brewing strains [28,40]. The altered carbohydrate-binding properties due to intergenic recombination could therefore have significant phenotypic consequences. These analyses revealed extensive recombination between the N-terminal domains (P -value $< 10^{-16}$), especially in the *FLO1*-like N-terminal domains. The reticulate with the N-terminal domains of the *FLO1*-like subgroup (Fig 2D) confirms several recombination events between different N-terminal *FLO1* gene domains (P -value $< 10^{-16}$).

The high frequency of recombination and lack of ancestral sequences or non-recombined outgroups prevent identification of the ancestral *FLO* sequences that recombined into the present-day alleles. Nevertheless, whereas it is impossible to discriminate between recombined and ancestral *FLO* genes, it is possible to identify groups of genes comprising the parents and the product of a recombination event, and to examine the recombination breakpoints (supplementary Fig S2 online). These analyses revealed several chimaeric adhesin sequences, including the previously described *LgFLO1* gene. Further, two types of recombination events appear to occur between adhesins, both with the potential to have strong phenotypic effect. In the first type, recombination events occur outside the repeat regions, across small regions of microhomology in the N- or C-terminal

domain (supplementary Fig S3 online). Detailed analysis of these genes revealed that many recombination events in the N-terminal domain occurred in-between regions important for substrate binding [35]. Such events could subtly alter the strength and preference of substrate binding in the N-terminal domain and therefore influence *FLO*. In the second group, recombination occurred across the central repeat domain. These recombination events lead to variation in the length and sequence of the repeats, which in turn can also result in new combinations of functional *FLO* domains.

Chimaeric adhesins also occur in pathogens

Previous studies in yeasts such as *Candida albicans* and *Candida glabrata* have speculated that the rapid phenotypic variation produced by chimaeric adhesins could contribute to pathogenicity in these species by avoiding the host immune response [39,41]. To investigate this, we performed a similar *in silico* analysis on adhesins of pathogenic yeasts by collecting sequences both for the *EPA* (epithelial adhesins) genes of *C. glabrata* (11 sequences) and the *ALS* (agglutinin-like sequences; 14 sequences) genes of *C. albicans*. Both cases revealed significant evidence for intergenic recombination (P -value $< 10^{-16}$), suggesting that recombination between adhesins is a common occurrence across yeast strains.

Engineered chimaeric adhesins confer distinct phenotypes

To assess the functional implications of recombination events between paralogs, we mimicked recombination by constructing several chimaeric *FLO* genes, each combining the 5' end of one

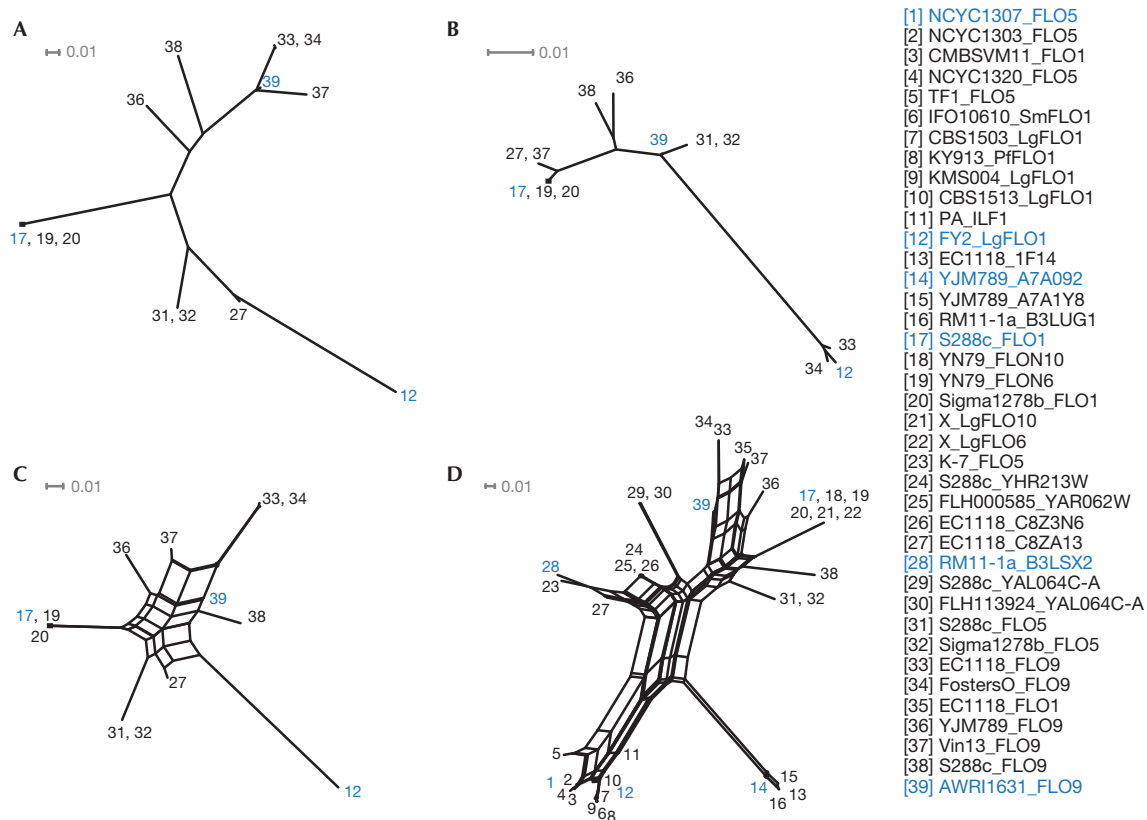


Fig 2 | Phylogenetic analyses of *FLO* genes reveal extensive intergenic recombination. A–C show the principle of reticulate analysis. Shown in blue are sequences used as representative haplotypes in supplementary Fig S2 online. (A) Phylogenetic tree of nucleotide sequences coding for N-terminal domains of all complete *FLO1*-like sequences. (B) Phylogenetic tree of nucleotide sequences coding for C-terminal domains of the same subset of *FLO1*-like sequences. (C) Trees of (A) and (B) are combined in an unrooted reticulate network (i.e., the overlay of trees shown in A and B). Such a phylogenetic reticulate allows to represent recombination events between either current or ancestral sequences and provides a more precise visualization of the evolutionary history of the *FLO* genes. Sequence (27) appears in different places in (A) and (B), thus placing it at an extension of the corner of a closed square in the reticulate tree. Such closed squares are generated by a predicted recombination event. (D) Reticulate displaying recombination events in the N-terminal domain of all *FLO1*-like sequences. Note that all these analyses were performed using all available *FLO* sequences but for the sake of clarity, only a subset of sequences was used to generate these figures. *FLO*, flocculation.

FLO gene with the 3' end of another (supplementary Table S3 online). We used three genes (*FLO1*, *FLO10* and *FLO11*) from S288c that represent *FLO* gene classes with distinct *FLO* characteristics as parents for the chimaeric genes. The protein products of the engineered chimaera retain the traditional three-domain structure of fungal adhesins while differing in N-terminal domains, total length, number and positions of tandem repeats and glycosylation sites.

To assess the functionality of these engineered chimaera, we expressed each construct separately in the non-flocculent and non-adherent S288c strain. Expression levels (determined by quantitative PCR) were comparable for each of the constructs, with an average expression level around 75% of the *ACT1* transcription level, which is comparable to the natural *FLO* gene expression in flocculent strains derived from the feral yeast EM93 [33].

We then determined flocculation strength, adhesion to diverse surfaces and cell wall hydrophobicity conferred by each of the chimaeric Flo proteins. Our results demonstrate that different

domains contribute to different phenotypes and that the chimaeric adhesins confer phenotypes that differ from their parental adhesins (Fig 3; Table 2; supplementary Table S4 online for details). This demonstrates that recombination events between paralogs can generate new alleles that can display new combinations of phenotypes and/or variation in the degree of adhesion.

CONCLUSION

Our results reveal an elegant mechanism by which gene duplicates might be used as a molecular toolbox for the generation of a large array of different new alleles and phenotypes. These findings also demonstrate that paralogs do not necessarily evolve completely independently through mutation. Instead, ectopic recombination events also contribute to the generation of sequence divergence in paralogs, which in turn propels evolutionary innovation [1–8].

Intergenic recombination events are observed predominantly in subtelomeric gene families, which are known to harbour

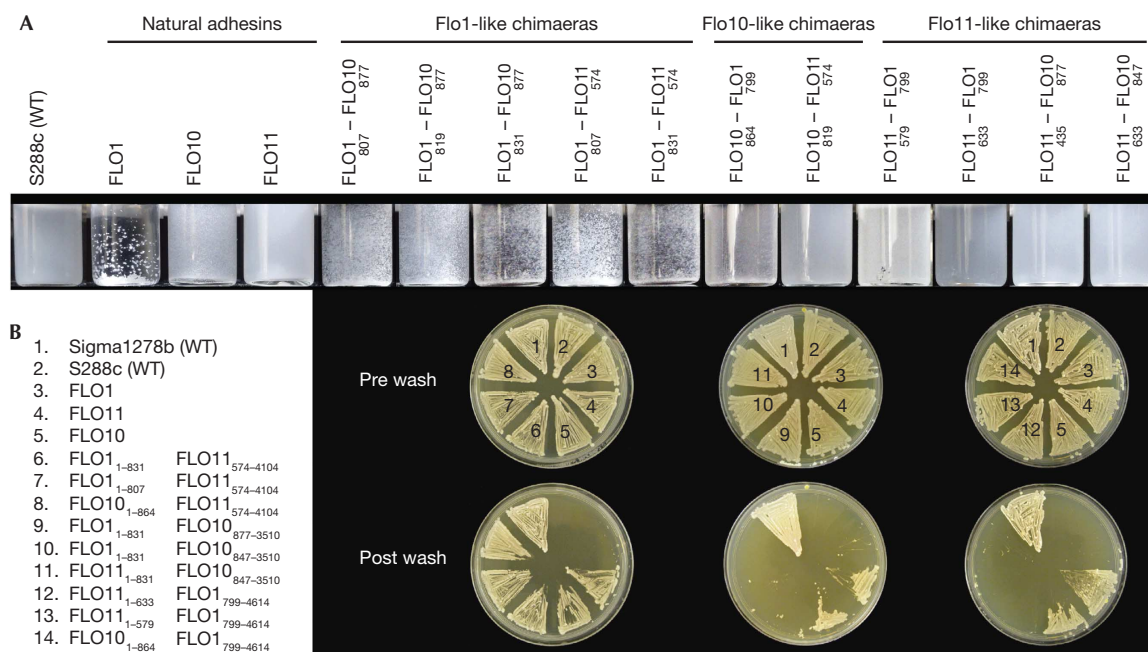


Fig 3 | Chimaeric adhesin confer new cell-cell adhesion phenotypes. To investigate the *FLO* (cell-cell adhesion) and agar-adhesion properties of chimaeric adhesins, three natural adhesin genes (*FLO1*, *FLO10* and *FLO11*) as well as several chimaeric adhesins consisting of the N-terminal part of one natural adhesin, and the C-terminal part of another were overexpressed in a strain that does not express any other adhesin gene (see Table 2; supplementary information online for details). (A) Ectopic recombination between *FLO* genes generates new chimaeric alleles that display a wide array of cell-cell adhesion (*FLO*) phenotypes. Strains expressing the natural *FLO1* adhesin or any chimaeric protein carrying the N-terminal part of *FLO1* all show strong *FLO*. Strains displaying adhesins with a *FLO10* N-terminal domain show a broad range of *FLO* that depends on the nature of the central- and C-terminal domain, whereas adhesins with a *FLO11* N-terminal domain confer weak or no *FLO* (for more information about the chimaeric adhesins see the supplementary Table S3 online). (B) The cell-surface adhesion of natural and chimaeric adhesin genes was measured by expressing these genes in cells that do not express any other adhesin gene. The resulting transformants were grown for 6 days and subsequently washed under a gentle stream of water to estimate their propensity to stick to the agar surface. Strains displaying adhesins containing the central and C-terminal part of *FLO11* show strong adhesion and therefore resist washing with water (see methods for details), whereas expression of other adhesins show a wide array of intermediate or weak cell-surface adhesion as measured with the plate-washing assay. *FLO*, flocculation; WT, wild-type.

lifestyle-specific genes, or in cell-surface gene families, which are involved in interactions with the cells environment. Hence, this mechanism could provide yeasts with an ever-changing reservoir of genes to quickly adapt and tune the way they interact with specific conditions and opportunities. Our results also confirm speculation that *C. albicans* and *C. glabrata* might have chimaeric adhesins [39,42], suggesting that this allows pathogens to adapt to the host. This mechanism is also similar to those of some pathogenic protozoans, such as *Plasmodium* and *Trypanosoma* spp. where modular cell coat proteins continuously recombine to form new variants and avoid host immune system recognition [16,17,43]. Hence, recombination of cell-surface and subtelomeric genes seems to be a common theme in (eukaryotic) microorganisms, yielding a substantial source of phenotypic variability with only a few genes. Moreover, as recombinations are clearly not limited to yeasts but have also been observed in the MHC class genes of vertebrates, it seems likely that similar mechanisms also exist in higher eukaryotes, including humans [20].

Our results show that in yeast, two distinct categories of gene families show enrichment for chimaeric alleles: families containing genes located near subtelomeres, and gene families encoding

cell-surface proteins. Importantly, this enrichment remains significant for both categories, even if all families belonging to the other category are removed from the analysis, indicating that the observed enrichment is independent. One main unresolved question is whether intergenic recombination occurs more frequently between subtelomeric and cell-surface genes, or whether all paralog families show similar rates of recombination but show differences in the adaptive advantage of such recombination events. It is, for example, possible that ectopic recombination between non-subtelomeric and non-cell-surface genes often leads to dysfunctional chimaera. Although it seems likely that both mechanisms might contribute to the enrichment of chimaeric alleles of cell-surface and subtelomeric genes, further research is needed to investigate the underlying mechanism(s).

METHODS

Paralog gene families were defined on the basis of relative pairwise blast scores between all known S288c proteins using the MCL algorithm (supplementary Dataset S1 online) [21]. For each family, the S288c sequences it comprises were blasted against a database containing 24 high-quality *Saccharomyces* genomes (supplementary Table S1 online). Resulting sequences were

Table 2 | Overview of adhesin phenotypes

<i>FLO_x</i>	<i>FLO_y</i>	Length	Floc.	Hydrop.	Agar A.	Polys A.
WT			11 ± 0	9 ± 5	+	1.0 ± 0.2
<i>FLO1</i>	—	4,614	98 ± 1	72 ± 6	+	3.0 ± 0.3
<i>FLO10</i>	—	3,510	55 ± 2	50 ± 10	+++	2.1 ± 0.3
<i>FLO11</i>	—	4,104	12 ± 5	80 ± 8	+++	1.5 ± 0.1
<i>FLO1</i> ₁₋₈₀₇	<i>FLO10</i> ₈₄₇₋₃₅₁₀	3,471	89 ± 6	61 ± 29	++	4.0 ± 0.3
<i>FLO1</i> ₁₋₈₀₇	<i>FLO10</i> ₈₇₇₋₃₅₁₀	3,441	59 ± 7	63 ± 9	++	4.3 ± 0.3
<i>FLO1</i> ₁₋₈₁₉	<i>FLO10</i> ₈₇₇₋₃₅₁₀	3,453	59 ± 6	80 ± 11	++	3.9 ± 0.2
<i>FLO1</i> ₁₋₈₃₁	<i>FLO10</i> ₈₇₇₋₃₅₁₀	3,465	78 ± 3	87 ± 2	++	4.0 ± 0.3
<i>FLO1</i> ₁₋₈₀₇	<i>FLO11</i> ₅₇₄₋₄₁₀₄	4,338	88 ± 4	69 ± 23	++++	4.5 ± 0.2
<i>FLO1</i> ₁₋₈₃₁	<i>FLO11</i> ₅₇₄₋₄₁₀₄	4,362	87 ± 7	80 ± 3	++++	4.0 ± 0.4
<i>FLO10</i> ₁₋₈₆₄	<i>FLO1799-4614</i>	4,680	75 ± 5	64 ± 2	+	3.7 ± 0.2
<i>FLO10</i> ₁₋₈₆₄	<i>FLO11</i> ₅₇₄₋₄₁₀₄	4,395	50 ± 4	65 ± 13	++++	1.2 ± 0.1
<i>FLO11</i> ₁₋₅₇₉	<i>FLO1799-4614</i>	4,395	14 ± 6	34 ± 6	+	0.8 ± 0.1
<i>FLO11</i> ₁₋₆₃₃	<i>FLO1799-4614</i>	4,449	18 ± 1	72 ± 8	+	0.7 ± 0.1
<i>FLO11</i> ₁₋₄₃₅	<i>FLO10</i> ₈₇₇₋₃₅₁₀	3,069	17 ± 2	54 ± 13	++	0.7 ± 0.1
<i>FLO11</i> ₁₋₆₃₃	<i>FLO10</i> ₈₄₇₋₃₅₁₀	3,297	12 ± 1	83 ± 6	++	0.9 ± 0.2

FLO, flocculation; WT, wild type. *FLO_x* and *FLO_y* represent the origin of the N-terminal and C-terminal part of the adhesin gene, respectively (with the numbers denoting the nucleotide positions in the open reading frame). Length is the length of the final adhesin. Floc. is the flocculation (%) of the strains expressing the adhesins. Hydrop. is the hydrophobicity level of a strain expressing the respective adhesin (%). Agar A. represents after the observed cell-surface adhesion as estimated by a plate-washing assay (+, weak adhesive growth; +++, strong adhesive growth). Polys A. represents the adherence of the strains to polystyrene relative to WT.

retained if they were not 200 bp shorter than the shortest member of the family and were at least 600 bp long. More (partial) coding sequences for *FLO* genes in the *Saccharomyces sensu stricto* group were gathered from the publicly available NCBI nucleotide database and the ENA database of EBI (supplementary Table S2 online). Sequences in a family were aligned using the ClustalW server provided by EBI [44]. For the *FLO* genes, N- and C-terminal sequence fragments were aligned separately. The alignments were analysed using SplitsTree4 [22]. We chose to ignore gaps and used the neighbour-joining algorithm to produce the trees [45]. The reticulate was generated using the NetworkNet algorithm [46]. Tests for significant evidence of recombination were performed using the PHI test provided by SplitsTree4 [46]. To check for statistically significant differences between the distributions of the PHI values, we performed a Komolgorov–Smirnov test in R. All yeast strains used in this study are listed in supplementary Table S5 online with primers used for construction, verification and quantitative PCR listed in supplementary Table S6 online. Culturing conditions for different experiments are described in supplementary information online. Strains and primers used for constructing reshuffled *FLO* genes are listed in supplementary Table S6 online. *FLO* tests were performed using the method of D'Hautcourt and Smart [47] with some modifications. Adhesive and invasive growth on agar were assessed by the plate-washing assay [48]. Adhesion to plastics was tested in 96-well flat-bottom plates (Microtest™ Flat bottom 96-well plate, Becton Dickinson Labware) according to Reynolds and Fink [29] with minor modifications. Hydrophobicity of cells expressing the different

(chimaeric) adhesins was measured by an adaptation of the method of Rosenberg [49].

Supplementary information is available at EMBO reports online (<http://www.emboreports.org>).

ACKNOWLEDGEMENTS

The authors thank Verstrepen Lab and Centre of Microbial and Plant Genetics members for valuable advice and feedback. This research was supported by ERC Young Investigator grant 241426, KU Leuven IDO program, the Belgian Federal Science Policy Office and European Space Agency PRODEX program and the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen, Belgium). KJV acknowledges support from ERC Stg 241426, VIB, KU Leuven, FWO Vlaanderen, the Odysseus program, EMBO YIP program and the AB InBev Baillet-Latour foundation.

Author contributions: J.F.C. and S.E.V.M. designed, performed and analysed (*in silico*) experiments and wrote the manuscript. J.D. and C.A.B. performed *in silico* analyses and wrote the manuscript. M.G.G. performed experiments. L.D.M., J.M., T.W., K.V. helped with design and interpretation of results and wrote the manuscript. K.J.V. led the study, designed and interpreted experiments and wrote the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* **9**: 938–950
- Ohno S (1970) Evolution by Gene Duplication (Springer-Verlag, Berlin).
- Hahn MW (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered* **100**: 605–617

4. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155
5. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV (2002) Selection in the evolution of gene duplications. *Genome Biol* **3**: RESEARCH0008
6. Hughes AL (2004) Birth-and-death evolution of protein-coding regions and concerted evolution of non-coding regions in the multi-component genomes of nanoviruses. *Mol Phylogenet Evol* **30**: 287–294
7. Talarico S, Whitefield SE, Fero J, Haas R, Salama NR (2012) Regulation of *Helicobacter pylori* adherence by gene conversion. *Mol Microbiol* **84**: 1050–1061
8. Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ (2012) Reconstruction of ancestral maltase enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. (in the press)
9. Brown CA, Murray AW, Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr Biol* **20**: 895–903
10. Fairhead C, Dujon B (2006) Structure of *Kluyveromyces lactis* subtelomeres: duplications and gene content. *FEMS Yeast Res* **6**: 428–441
11. Linardopoulou E, Mefford HC, Nguyen O, Friedman C, van den Engh G, Farwell DG, Coltrera M, Trask BJ (2001) Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location. *Hum Mol Genet* **10**: 2373–2383
12. Louis EJ, Naumova ES, Lee A, Naumov G, Haber JE (1994) The chromosome end in yeast: its mosaic nature and influence on recombinational dynamics. *Genetics* **136**: 789–802
13. Verstrepen KJ, Fink GR (2009) Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi. *Annu Rev Genet* **43**: 1–24
14. Barry JD, Ginger ML, Burton P, McCulloch R (2003) Why are parasite contingency genes often associated with telomeres? *Int J Parasitol* **33**: 29–45
15. Berriman M et al (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**: 416–422
16. Taylor JE, Rudenko G (2006) Switching trypanosome coats: what's in the wardrobe? *Trends Genet* **22**: 614–620
17. Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, Nehrbass U, Wellemes TE, Scheref A (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**: 1018–1022
18. Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, Peterson DS, Pinches R, Newbold CI, Miller LH (1995) Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**: 101–110
19. Cramer A, Raillard SA, Bermudez E, Stemmer WP (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**: 288–291
20. Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* **29**: 217–222
21. Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**: 1575–1584
22. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267
23. Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681
24. Verstrepen KJ, Klis FM (2006) Flocculation, adhesion and biofilm formation in yeasts. *Mol Microbiol* **60**: 5–15
25. Dranginis AM, Rauceo JM, Coronado JE, Lipke PN (2007) A biochemical guide to yeast adhesins: glycoproteins for social and antisocial occasions. *Microbiol Mol Biol Rev* **71**: 282–294
26. Hoyer LL (2001) The ALS gene family of *Candida albicans*. *Trends Microbiol* **9**: 176–180
27. Verstrepen KJ, Reynolds TB, Fink GR (2004) Origins of variation in the fungal cell surface. *Nat Rev Microbiol* **2**: 533–540
28. Govender P, Domingo JL, Bester MC, Pretorius IS, Bauer FF (2008) Controlled expression of the dominant flocculation genes FLO1, FLO5, and FLO11 in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **74**: 6041–6052
29. Reynolds TB, Fink GR (2001) Bakers' yeast, a model for fungal biofilm formation. *Science* **291**: 878–881
30. Stovicek V, Vachova L, Kuthan M, Palkova Z (2010) General factors important for the formation of structured biofilm-like yeast colonies. *Fungal Genet Biol* **47**: 1012–1022
31. Van Mulders SE, Christianen E, Saerens SM, Daenen L, Verbelen PJ, Willaert R, Willuert R, Verstrepen KJ, Delvaux FR (2009) Phenotypic diversity of Flo protein family-mediated adhesion in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **9**: 178–190
32. Stratford M (1992) Yeast flocculation: a new perspective. *Adv Microb Physiol* **33**: 2–71
33. Smukalla S et al (2008) FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell* **135**: 726–737
34. Zupancic ML, Frieman M, Smith D, Alvarez RA, Cummings RD, Cormack BP (2008) Glycan microarray analysis of *Candida glabrata* adhesin ligand specificity. *Mol Microbiol* **68**: 547–559
35. Veelders M, Bruckner S, Ott D, Unverzagt C, Mosch HU, Essen LO (2010) Structural basis of flocculin-mediated social behavior in yeast. *Proc Natl Acad Sci USA* **107**: 22511–22516
36. Verstrepen KJ, Jansen A, Lewitter F, Fink GR (2005) Intragenic tandem repeats generate functional variability. *Nat Genet* **37**: 986–990
37. Fidalgo M, Barrales RR, Jimenez J (2008) Coding repeat instability in the FLO11 gene of *Saccharomyces* yeasts. *Yeast* **25**: 879–889
38. Cormack BP, Ghori N, Falkow S (1999) An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells. *Science* **285**: 578–582
39. Butler G et al (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* **459**: 657–662
40. Kobayashi O, Hayashi N, Kuroki R, Sone H (1998) Region of FLO1 proteins responsible for sugar recognition. *J Bacteriol* **180**: 6503–6510
41. Zhao X, Oh SH, Jajko R, Diekema DJ, Pfaller MA, Pujol C, Soll DR, Hoyer LL (2007) Analysis of ALS5 and ALS6 allelic variability in a geographically diverse collection of *Candida albicans* isolates. *Fungal Genet Biol* **44**: 1298–1309
42. De Las Penas A, Pan SJ, Castano I, Alder J, Cregg R, Cormack BP (2003) Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to RAP1- and SIR-dependent transcriptional silencing. *Genes Dev* **17**: 2245–2258
43. Roth C, Bringaud F, Layden RE, Baltz T, Eisen H (1989) Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. *Proc Natl Acad Sci USA* **86**: 9375–9379
44. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680
45. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–25
46. Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**: 255–265
47. D'Hautcourt O, Smart K (1999) Measurement of brewing yeast flocculation. *J Am Soc Brew Chem* **57**: 123–128
48. Roberts RL, Fink GR (1994) Elements of a single MAP kinase cascade in *Saccharomyces cerevisiae* mediate two developmental programs in the same cell type: mating and invasive growth. *Genes Dev* **8**: 2974–2985
49. Rosenberg M (1984) Bacterial adherence to hydrocarbons: a useful technique for studying cell surface hydrophobicity. *FEMS Microbiol Lett* **22**: 289–295



EMBO reports is published by Nature Publishing Group on behalf of European Molecular Biology Organization. This article is licensed under a Creative Commons Attribution Noncommercial No Derivative Works 3.0 Unported License [<http://creativecommons.org/licenses/by-nc-nd/3.0>]