Chromosome Number Reduction in *Eremothecium coryli* by Two Telomere-to-Telomere Fusions

Jürgen Wendland* and Andrea Walther

Carlsberg Laboratory, Yeast Biology, Copenhagen, Denmark

*Corresponding author: E-mail: juergen.wendland@carlsberglab.dk.

Abstract

The genus *Eremothecium* belongs to the *Saccharomyces* complex of pre-whole-genome duplication (WGD) yeasts and contains both dimorphic and filamentous species. We established the 9.1-Mb draft genome of *Eremothecium coryli*, which encodes 4,682 genes, 186 tRNA genes, and harbors several Ty3 transposons as well as more than 60 remnants of transposition events (LTRs). The initial de novo assembly resulted in 19 scaffolds, which were assembled based on synteny to other *Eremothecium* genomes into six chromosomes. Interestingly, we identified eight *E. coryli* loci that bear centromeres in the closely related species *E. cymbalariae*. Two of these *E. coryli* loci, *CEN1* and *CEN8*, however, lack conserved DNA elements and did not convey centromere function in a plasmid stability assay. Correspondingly, using a comparative genomics approach we identified two telomere-to-telomere fusion events in *E. coryli* as the cause of chromosome number reduction from eight to six chromosomes. Finally, with the genome sequences of *E. coryli*, *E. cymbalariae, and Ashbya gossypii* a reconstruction of three complete chromosomes of an *Eremothecium* ancestor revealed that *E. coryli* is more syntenic to this ancestor than the other *Eremothecium* species.

Key words: *Saccharomyces*, whole-genome sequencing, genome evolution, ancestral gene order, centromere DNA elements, synteny, paleogenomics.

Introduction

Comparative genomics is most powerful when comparing essentially complete draft genomes. This can yield insight into the evolution of species and compiling several genomes of closely related species may allow the reconstruction of ancestral genomes. The precision of such a paleogenomic reconstruction depends on the degree of synteny, that is, conserved gene order in the studied species and on the number of sequenced genomes (Bhutkar et al. 2007; Muffato and Roest Crollius 2008; El-Mabrouk and Sankoff 2012).

Yeast species of the *Saccharomyces* complex have been of considerable interest based on their fermentative properties and their large evolutionary timescale spanning at least 100 Ma from an ancient whole-genome duplication (WGD) event (Wolfe and Shields 1997). Compiling the data of 11 sequenced yeast species a pre-WGD ancestor was reconstructed harboring 4,700 genes distributed on eight chromosomes (Gordon et al. 2009). Due to a WGD modern *Saccharomyces* sensu stricto species contain 16 chromosomes per haploid genome. From an ancestral genome, the evolutionary paths in terms of duplications, inversions, and reciprocal translocations can be inferred. Interestingly, a comparison of the protoploid *Lachancea kluyveri*, which contains eight chromosomes, with this pre-WGD ancestor allowed the reconstruction of the complete evolutionary genome rearrangement history of *L. kluyveri* (Gordon et al. 2011). Chromosome number, however, is not static and several protoploid, that is, "pre-WGD" and post-WGD species of the *Saccharomyces* complex have undergone chromosome number reductions.

There are basically two mechanisms for a reduction in chromosome number without loss of coding information: 1) By telomere-to-telomere fusion and inactivation of one of the two centromeres of such a newly formed chromosome or 2) by breakage of a chromosome at a centromere and fusion of the two chromosomal arms to two telomeres of other chromosomes. The first seems to be more widespread than the latter as breakage of a chromosome at a centromere was so far only observed in *Eremothecium/Ashbya gossypii* (Gordon et al. 2011).

The genus *Eremothecium* constitutes clade 12 of the *Saccharomyces* complex (Kurtzman and Robnett 2003). The

type of strain of this genus, *Eremothecium cymbalariae*, was first isolated and described in 1888 by Borzi and recently its genome sequence has been determined (Borzi 1888; Wendland and Walther 2011). *Eremothecium* species are known to cause fruit rotting, for example, on cotton or tomato (Miyao et al. 2000). Insect vectors are required for dispersal of the fungi, particularly milkweed bugs, boxelder bugs, or other stink bugs (Dietrich et al. 2013). The disease caused is referred to as stigmatomycosis or "yeast spot disease" (Ashby and Nowell 1926).

Major interest in *Eremothecium* species was attracted by *A. gossypii* as a potent overproducer of riboflavin/vitamin B$_2$ (Kato and Park 2012). Based on its molecular genetic tractability, *Ashbya* soon became a model for studies of fungal cell biology and filamentous growth (Wendland and Walther 2005). Comparisons of the complete genomes of the filamentous fungi *A. gossypii* and *E. cymbalariae* revealed that *E. cymbalariae* harbors greater similarity to the pre-WGD ancestor than *A. gossypii* (Dietrich et al. 2004; Wendland and Walther 2011). This includes 1) eight chromosomes in *E. cymbalariae* compared with only seven in *A. gossypii*, 2) a low GC content of 40.3% in *E. cymbalariae* (as found in other yeast species) versus the remarkably high GC content of 51.8% in *A. gossypii*, 3) larger blocks of synteny, 4) a similar gene density between *E. cymbalariae* and the yeast ancestor, and 5) the presence of a Ty3 transposon in *E. cymbalariae*, which is absent in *A. gossypii* (Wendland and Walther 2011). *Ashbya gossypii* is thus characterized by a more divergent, more rearranged, and much more compact genome—largely due to size reductions in intergenic regions—compared with the *E. cymbalariae* genome.

The *Eremothecium* genus is not only composed of true filamentous fungi but it contains also dimorphic yeasts, for example, *Nematospora/Holleya sinecauda* and *Nematospora/Eremothecium coryli*. Although *E. cymbalariae* and *A. gossypii* grow only in the filamentous form, dimorphic fungi generate yeast cells, pseudohyphal cells, or filaments. Emil Christian Hansen, who worked at the Carlsberg Laboratory, first described the genus *Nematospora* in 1904 (Hansen 1904). Later *Ashbya*, *Nematospora*, *Holleya*, and *Eremothecium* were placed in a single genus that was seeded within the *Saccharomycetaceae* (Kurtzman 1995; Prillinger et al. 1997). This grouping suggested that filamentous growth may have been gained in the *Eremothecium* genus whereas the yeast ancestor was unicellular/dimorphic (Schmitz and Philippsen 2011). To further elucidate genome evolution in *Eremothecium*, we established the draft genome of the dimorphic species *E. coryli*. Using comparative genomics and functional analysis tools, we identified the mechanism of chromosome number reduction from 8 to 6 chromosomes in *E. coryli*. Furthermore, based on conserved synteny, three chromosomes of an *Eremothecium* ancestor (ERA) could be reconstructed. Comparisons of the recent *Eremothecium* genomes with ERA indicate that *E. coryli* is most syntenic to ERA

supporting the hypothesis that the lineage ancestor was a unicellular/dimorphic yeast and true filamentous growth may be an apomorphy in the *Eremothecium* lineage.

Materials and Methods

Strains and Media

Eremothecium coryli strain CBS 5749 was sequenced. For plasmid stability assays *H. sinecauda* (CBS 8199) served as a host. Strains were grown using complete media (1% yeast extract, 1% peptone, and 2% dextrose) supplemented with G418/geneticin (200 μg/ml) for the selection of antibiotic-resistant plasmid transformants or minimal media with either asparagine or ammonium sulfate as nitrogen source. For plasmid propagation, *Escherichia coli* DH5α was used.

Transformation of *H. sinecauda*

Transformation and plasmid stability assays in *H. sinecauda* were done as described previously (Schade et al. 2003).

Plasmid Constructs

Episomal plasmids were generated for testing of plasmid stability and centromere activity. To this end centromere DNA fragments of the *E. coryli* centromere loci of chromosome 1 (734 bp), 2 (1,075 bp), 3 (785 bp), 4 (821 bp), 7 (772 bp), and 8 (445 bp) were amplified by polymerase chain reaction and cloned into the high copy (autonomously replicating sequence [ARS]-containing) shuttle vector pHC shuttle (#310; Schade et al. 2003) using *Xba*I and *Xho*I restriction sites provided with the primers. This generated plasmids C875-C880. A low copy pLC shuttle (#268) containing *A. gossypii* ARS and centromere DNA sequences was used as a control.

Sequencing Strategy

The *E. coryli* genome was sequenced using Illumina HiSeq2000 next-generation sequencing with 100-bp paired-end reads and an 8-kb mate-pair library (LGC Genomics, Berlin, Germany). Sequencing generated approximately 40 million reads corresponding to more than 100× coverage of the *E. coryli* genome. Assembly of the genome sequencing data produced 19 scaffolds/supercontigs.

Annotation of the *E. coryli* Genome

The 19 scaffolds of the *E. coryli* draft genome were submitted to GenBank with a BioProject number (PRJNA229863) and have been deposited under accession number AZAH00000000. The mitochondrial genome has not been assembled.

The *E. coryli* genes were compared with the *A. gossypii*, *E. cymbalariae*, and *Saccharomyces cerevisiae* genomes available from *Ashbya* Genome Database (http://agd.vital-it.ch/index.html, last accessed May 15, 2014) and *Saccharomyces*

Genome Database (http://www.yeastgenome.org, last accessed May 15, 2014) and GenBank using local blast tools (available at http://blast.ncbi.nlm.nih.gov, last accessed May 15, 2014). LTR sequences were identified using BLASTN. Fine annotation of the *E. coryli* genome used syntenic relationships to *A. gossypii*, *E. cymbalariae*, and *S. cerevisiae*. Unidentified *E. coryli* ORFs were also searched against the nonredundant data set of National Center for Biotechnology Information. The assembly of the *E. coryli* genome into six chromosomes was based on syntenic gene order and the prediction of reciprocal translocations. A systematic nomenclature based on this chromosome assembly was generated. As species identifier for *E. coryli* "Eco_" was used followed by the chromosome number (1.–6.) and the feature number (1–*n* starting from the first ORF at the left telomere running continuously to the last ORF [*n*] at the right telomere of the chromosome, e.g., Eco_1.001 for the first ORF at the left end of chromosome 1). For the identification of tRNA genes, tRNAscan (http://lowelab.ucsc.edu/tRNAscan-SE/, last accessed May 15, 2014) was used (Schattner et al. 2005).

Results

Eremothecium Genome Comparisons

Eremothecium coryli is a dimorphic fungus that lacks dichotomous tip branching characteristic for hyphal tip growth in its filamentous relatives *A. gossypii* and *E. cymbalariae* (Gastmann et al. 2007). The *E. coryli* strain CBS 5749 was sequenced using Illumina HiSeq2000 with 8 kb mate-pair libraries and paired-end sequencing with more than 100× genome coverage. The draft genome was assembled into 19 scaffolds (table 1). The genome size is approximately 9.1 Mb and thus of intermediate size compared with *E. cymbalariae* (9.7 Mb) and *A. gossypii* (8.7 Mb). We identified 4,682 genes, which is close to the slightly over the 4,700 genes for the other *Eremothecium* species indicating that our assembly is basically complete. The *E. coryli* genome consists of 73.6% encoding DNA with a GC content of 41.5% very similar to *E. cymbalariae* (73.6% coding with 40.3% GC) and in contrast to *A. gossypii* (79.5% coding and 51.8% GC). The apparently higher similarity between the *E. coryli* and *E. cymbalariae* genomes is also reflected by the amount of synteny blocks: Longer stretches of conserved gene order between these two species result in fewer synteny blocks (139) compared with *E. coryli* and *A. gossypii* (198) (see table 1). Interestingly, we also identified several Ty3 transposons and 83 remnants of transposition marked by LTRs (supplementary table S1, Supplementary Material online). Of these LTRs 73, that is 88%, are adjacent to tRNA genes in *E. coryli* (supplementary table S4, Supplementary Material online). The paired-end sequencing and scaffold assembly indicate that there are at least six full-length Ty3 transposons present in the *E. coryli* genome. Sequence analysis of the *E. cymbalariae* genome

indicated only one Ty3 transposon that—based on the orientation of the LTRs—may, however, have lost its ability to transpose. We also found several LTRs positioned at the end of scaffolds in *E. coryli*. In three cases, we inferred reciprocal translocations at these positions for the assembly of the *E. coryli* genome (see below).

Morphological differences between the filamentous *Eremothecium* species *E. cymbalariae* and *A. gossypii* compared with the dimorphic species including *H. sinecauda* and *E. coryli* are not necessarily also manifested in the average similarity of the protein-coding genes. Comparison of the proteomes between the three sequenced species shows an average identity of approximately 60% between these species, which is slightly higher between *E. coryli* and *E. cymbalariae* (63.2%) compared with *E. coryli* and *A. gossypii* (62.3%) (fig. 1*A*). Overall the three *Eremothecium* species share about 95% of their genes. Furthermore, *E. coryli* shares an additional 1% of its genes with *E. cymbalariae* but not with *A. gossypii* and a similar number with *A. gossypii* but not with *E. cymbalariae* (fig. 1*B*).

Eremothecium species are pre-WGD and thus contain unduplicated protoploid genomes. Yet, these species are not completely devoid of gene duplications. Some of them occur dispersed throughout the genome but others are present as tandem duplications. These give rise to evolutionary diversification and subfunctionalization as has been demonstrated for *RHO1* paralogs in *A. gossypii* (Walther and Wendland 2005; Köhli et al. 2008). Out of 21 tandem duplications found in *A. gossypii*, *E. coryli* shares 13 and *E. cymbalariae* 9 (supplementary table S2, Supplementary Material online). The remaining *A. gossypii* duplications are either telomeric in *A. gossypii* or may hint to species-specific functions, for example, *A. gossypii MCH4*, which is currently under investigation. In addition to these shared duplications, there are seven tandem duplications that are specific for *E. coryli*. Interestingly, *ABR156W/YJL212C* occurs in four tandem copies. *YJL212C* encodes the oligopeptide transporter *OPT1* in *S. cerevisiae*, which also transports phytochelatin (Osawa et al. 2006). This multiplication may be functionally relevant for metal homeostasis. Furthermore, there is a tandem duplication of the *E. coryli* paralogs of *AER22W/YBR139W*, which encodes a serine carboxypeptidase that is required for phytochelatin synthesis in yeast (Wünschmann et al. 2007). This suggests a functional linkage of these duplications that is specific for *E. coryli*.

Synteny Relationships within *Eremothecium* Species

Synteny describes the conservation of gene order and transcriptional orientation of homologous genes between two-related species. Comparisons of the *E. coryli* genome with those of *E. cymbalariae* and *A. gossypii* revealed four types of synteny relationships (fig. 2). First, by far the largest parts of all three *Eremothecium* genomes show synteny

Table 1

Eremothecium coryli Genome Summary

Scaffold[a]	Number of Genes	Scaffold Length (bp)	% Encoding	GC Content (%)	tRNAs	LTRs[b]	Blocks to *Eremothecium cymbalariae*[c]	Blocks to *Ashbya gossypii*
0	1,012	1,827,054	76.0	41.59	27	5 (1)	26	47
1	567	1,105,492	75.7	41.34	33	13 (1)	19	25
2	536	1,035,239	73.9	41.39	8	3	12	20
3	521	1,037,803	73.2	41.04	30	11 (2)	19	26
4	277	590,229	72.6	41.11	9	6	6	7
5	275	544,843	75.7	41.59	10	5	14	16
6	254	521,725	72.9	41.37	8	2	2	9
7	251	503,936	70.0	41.10	9	1	13	13
8	235	488,308	72.9	40.74	7	3	6	7
9	217	415,668	71.7	41.69	6	2	6	10
10	143	262,922	75.7	41.30	13	4	3	5
11	113	222,772	69.8	40.98	7	2	6	5
12	98	173,376	71.6	42.47	2	0	2	3
14	44	85,504	68.2	39.71	2	0	2	2
15+13	69	142,227	69.4	40.29	8	5 (1)	1	1
16	30	53,536	60.6	40.82	1	0	1	1
17+18	40	84,386	59.1	42.13	7	3 (1)	1	1
	4.682	9.095.020	73.6	41.57	187	65	139	198

[a]Scaffolds 15+13 and 17+18 were combined based on synteny.
[b]LTRs were identified based on the direct repeat sequences flanking full-length Ty3 transposons (number in brackets).
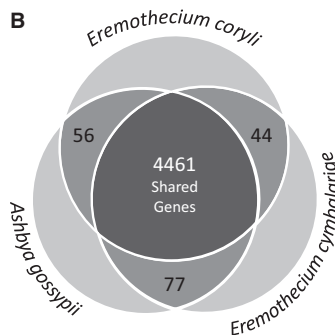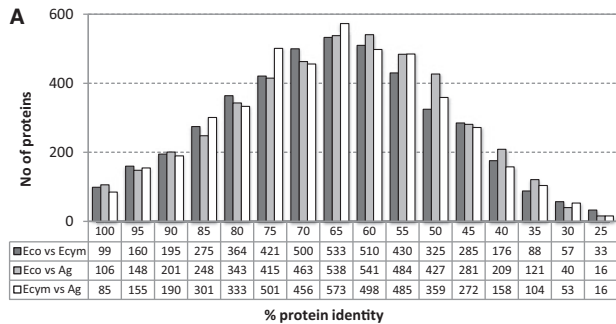[c]Block synteny based on conserved gene order.

FIG. 1.—Proteome and genome comparisons. (*A*) Pairwise proteome comparisons between *Eremothecium coryli*, *E. cymbalariae*, and *Ashbya gossypii* using all protein-coding genes of these *Eremothecium* species. (*B*) Diagram showing the distribution of homologous genes within *Eremothecium* species. Central genes (4,461 of ~4,700) are shared by all three species. Genes in intersections are shared by only two species.

between all *Eremothecium* species. A long stretch of conserved synteny encompassing, for example, 108 genes or 230 kb of DNA, is found at the centromere locus of *E. coryli* chromosome 6 (fig. 2*A*). Second, there are regions of single block synteny between *E. coryli* and *A. gossypii* that are fragmented into multiple blocks in the *E. cymbalariae* genome. One example of 44 genes distributed over 85 kb on *E. coryli* chromosome 3 is shown in figure 2*B* (see below for chromosome assignments). The syntenic *A. gossypii* locus harbors the genes from AAL174C to AAL131C. Homologs of these genes are found in five blocks on four different chromosomes in *E. cymbalariae* (fig. 2*B*). Conversely, there are regions of single block synteny between *E. coryli* and *E. cymbalariae* that are dispersed to multiple regions in the *A. gossypii* genome (fig. 2*C*). In the example shown, also derived from *E. coryli* chromosome 3, 78 genes found on 138 kb in *E. coryli* are syntenic to *E. cymbalariae* Ecym_5.451 to Ecym_5.528. Finally, there are positions in the *E. coryli* assembly in which both *A. gossypii* and *E. cymbalariae* genomes show synteny breaks. However, we found several locations in which the *E. coryli* gene order is syntenic with that of the pre-WGD ancestor (fig. 2*D*). The region of synteny shown harbors 106 genes on 205 kb dispersed on three to four chromosomes in *E. cymbalariae* and *A. gossypii*, respectively. An analysis of the *E. coryli* genome for positions of such conserved ancient synteny between *E. coryli* and the yeast ancestor that are not conserved in either *A. gossypii* or *E. cymbalariae* identified 20 such cases (supplementary table S3, Supplementary Material online). Eleven of
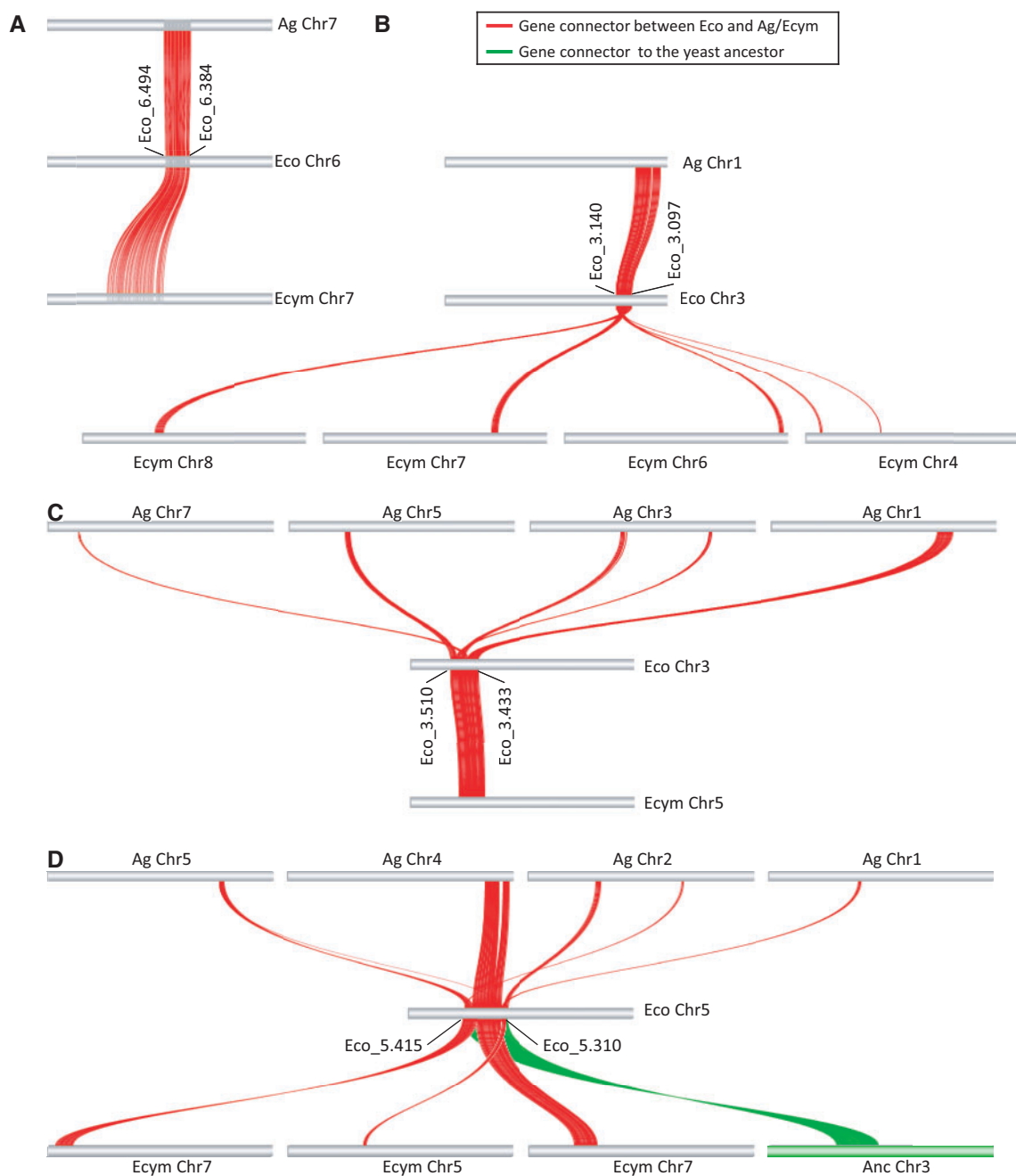
Fig. 2.—Synteny relationships in *Eremothecium* genomes. (*A*) Single block synteny among *Ashbya gossypii* (Ag), *Eremothecium coryli* (Eco), and *E. cymbalariae* (Ecym). See text and supplementary material, Supplementary Material online, for the *E. coryli* chromosome assignments and the *E. coryli* systematic gene nomenclature. (*B*) Single block synteny between *E. coryli* and *A. gossypii* but not between *E. coryli* and *E. cymbalariae*. (*C*) Single block synteny between *E. coryli* and *E. cymbalariae* but not between *E. coryli* and *A. gossypii*. (*D*) Conserved ancient synteny between *E. coryli* and the reconstructed pre-WGD ancestor (Anc) not found in *A. gossypii* and *E. cymbalariae*. Such cases not only support our scaffold assembly but are also instrumental in generating an ancestral gene order. Red connectors were used to link each homologous gene pair between *Eremothecium* species; green connectors in (*D*) were used to link homologs between *E. coryli* and the pre-WGD ancestor. Graphs were generated using Strudel software (http://bioinf.hutton.ac.uk/strudel/, last accessed May 15, 2014).

these were found to be associated with tRNA genes that often occur at breakpoints of synteny. All tRNAs and their scaffold positions are listed in supplementary table S4, Supplementary Material online. Due to the efficient homologous

recombination machinery in *Eremothecium*, short homology regions provided, for example, by tRNA genes can readily serve as templates for reciprocal translocations (Steiner et al. 1995). The examples presented in figure 2*B–D* indicate
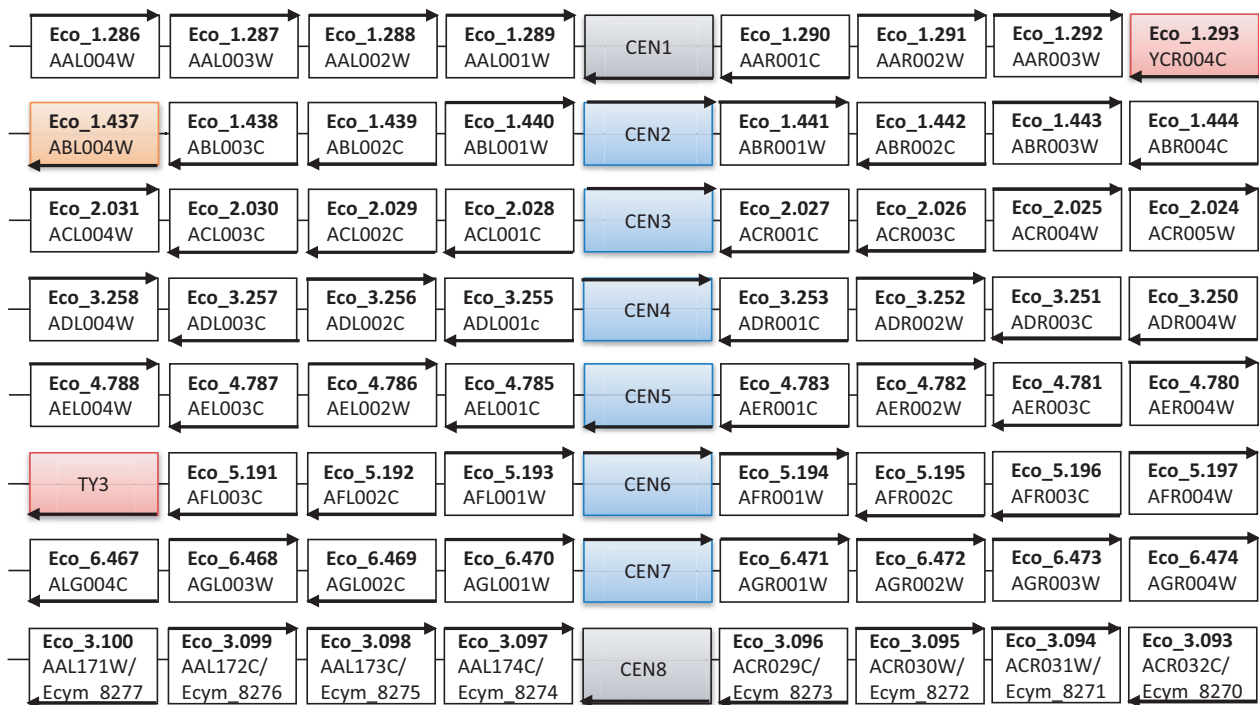
Fig. 3.—Centromere loci in *Eremothecium*. Identification of eight *Eremothecium coryli* loci harboring six functional centromeres was based on synteny to *Ashbya gossypii* and *E. cymbalariae*. Arrows indicate transcriptional orientation of genes. Arrows for centromeres indicate the orientation of centromere DNA elements (*CDEI–CDEII–CDEIII*). Special features are highlighted (*YCR004C* and TY3 absent from *A. gossypii* and *E. cymbalariae*; *ABL004W* absent from *E. cymbalariae*) and systematic gene nomenclature was used for each species. *Eremothecium coryli CEN1* and *CEN8* do not harbor conserved centromere DNA elements (see also fig. 4).

species-specific genome evolution events. Of course, they are by far outnumbered by syntenic gene organization. Yet, these regions could be drivers of species-specific evolution and thus of interest for targeted functional analyses.

Identification of Centromere Loci in *E. coryli* Scaffolds

Previously, we identified eight centromere loci in *E. cymbalariae* providing evidence that an ERA, similarly to the yeast ancestor, also contained eight chromosomes (Wendland and Walther 2011). By searching for homologs of centromere-associated *E. cymbalariae* genes in *E. coryli*, we identified all eight syntenic loci (fig. 3). At these loci, some additions are present in *E. coryli*, for example, a *YCR004C* homolog of unknown function that is absent from both *A. gossypii* and *E. cymbalariae*. These loci provide clear direction for the search for centromere DNA in *E. coryli*. Centromere DNA in *Eremothecium* is very similar to that of *S. cerevisiae* in that there are conserved centromere DNA elements (*CDEI*, *CDEII*, and *CDEIII*) with the sole difference that the AT-rich *CDEII* is twice as long in *Eremothecium* as in *S. cerevisiae* (Dietrich et al. 2004). Alignment of the putative centromere regions allowed the identification of six bona fide centromeres in *E. coryli*. In the syntenic *E. coryli* region harboring *CEN8* in *E. cymbalariae*, we could not locate any centromere DNA. For the syntenic

region of *CEN1* similarity to the core sequence of *CDEIII* was found, however, the surrounding sequence did not match the *CDEIII* consensus and, furthermore, *CDEI* was not present. Moreover, two of the centromere loci, *CEN4* and *CEN8*, are located on scaffold 1 (fig. 4). This suggests that only six of these eight loci harbor functional centromeres. To test for centromere function of the *E. coryli CEN1* and *CEN8* loci in vivo, we used a plasmid stability assay that was originally developed for yeast (Murray and Szostak 1983). *Holleya sinecauda/E. sinecaudum* served as a host as previously described (Schade et al. 2003). In this assay, transformants harboring *ARS*-plasmids will form only small colonies compared with transformants carrying *CEN-ARS*-plasmids, which is based on the improved segregation properties of centromere-bearing plasmids. Because of the plasmid-encoded antibiotic resistance gene, daughter cells without plasmid are sensitive to the antibiotic and die. With this assay, we could demonstrate that the intergenic regions of, for example, *CEN4* and *CEN7* harbor functional centromeres whereas *E. coryli CEN1* and *CEN8* are nonfunctional (fig. 5).

Chromosome Number Reduction in *E. coryli*

The previous section indicated that *E. coryli* has decommissioned two centromeres. As we identified eight syntenic

CEN	CDEI	CDEII	CDEIII	E. coryli scaffold	E.coryli CHR
E.coryli CEN1	TAGGG**ACC**GC	– 164 bp –	AGGAGCATCC**ATCCGAA**TGTATAAAGGTTAT	10	1
E.coryli CEN2	AT**CACCTG**	– 165 bp –	TGTGTTCGCT**ATCCGAA**CGTATATTATATTTT	11	1
E.coryli CEN3	AT**CACCTG**	– 165 bp –	TGTCTTAGTT**TTCCGAA**GAGATATTTTT	15	2
E.coryli CEN4	AT**CACCTG**	– 165 bp –	TGTATCATGC**TTCCGAA**CGTAGAAATAATTT	1	3
E.coryli CEN5	TAC**CACCTG**	– 166 bp –	TGTATAGTGC**TTCCGAA**CATAGTTTAAATTTT	0	4
E.coryli CEN6	TT**CACCTG**	– 164 bp –	TGTATAAGTT**TTCCGAA**CATAAATAATAATTTT	3	5
E.coryli CEN7	AT**CATCTG**	– 165 bp –	TGTATATCAG**TTCCGAA**CATATAAATAATAT	4	6
E.coryli CEN8	none	–	none	1	3

FIG. 4.—Analysis of centromere DNA elements in *Eremothecium coryli*. *CEN* sequences were identified based on the highlighted *CDEI* (CAYCTG) and *CDEIII* (TCCGAA) consensus sequences. The *CDEII* spacers are AT rich (>70%) and about 165 bp in length. The intergenic region between the *E. coryli* homologs of *AAL174C* and *ACR029C* is only 291 bp lacking conserved sequences for *CEN8 (marked as* CEN8*)*. *EcoCEN1* sequence is without conserved *CDEI* and with only partially conserved *CDEIII (*CEN1*)*. Positions of these loci on *E. coryli* scaffolds and assembled chromosomes (see below) are indicated.

centromere loci in *E. coryli*, this can be explained by two cases of telomere-to-telomere fusion of two chromosomes. Concomitant with each telomere-to-telomere fusion, loss of function mutations in one of the two centromeres of each new chromosome must have occurred. In total *E. coryli* should thus contain six chromosomes. We therefore analyzed the *E. coryli* genome data for traces of these telomere-to-telomere fusion events.

The reconstructed pre-WGD ancestor provides 8 chromosomes with 16 ancient telomeres (Gordon et al. 2009). Remarkably, 15 of these loci are conserved at telomeres in *E. cymbalariae* and 9 out of those loci are also at telomeres in *A. gossypii* (fig. 6). We then went on to identify the scaffold positions of the respective telomere-linked genes in *E. coryli*. Ten of these were located at scaffold ends, six were internal. Interestingly, two scaffolds, S5 and S7, harbor homologs located at two different telomeres in the pre-WGD ancestor each (fig. 6). Strikingly, these telomeric loci are directly adjacent to each other on both scaffolds providing direct evidence for two telomere-to-telomere fusion events. According to the nomenclature of the yeast ancestor, these fusions involved the telomeres of Anc3R and Anc8R in one case and Anc6R and Anc7L in the other (fig. 7A and B). Interestingly, the telomere-to-telomere fusion located on scaffold 5 would not have been detected unambiguously without the reconstructed pre-WGD ancestral genome. The respective homologs in *A. gossypii* are found at internal positions in three different chromosomes. In *E. cymbalariae*, the telomere of Anc_3R is also telomeric at chromosome 6L, whereas the telomere of the ancestral chromosome 8R became internalized.

Evidence of a telomere-to-telomere fusion found in *E. coryli* scaffold 7 is based both on conservation in *Eremothecium* and the pre-WGD ancestor. In *A. gossypii*, one telomeric end is conserved, whereas the location of *ACR293C* is telomeric both
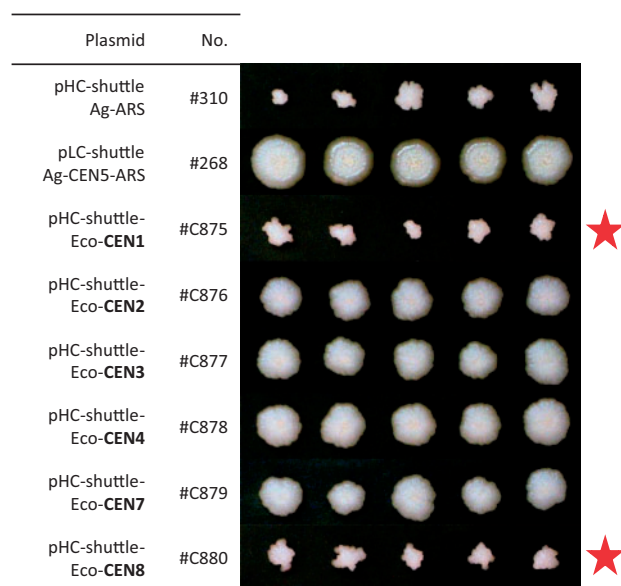
Plasmid	No.
pHC-shuttle Ag-ARS	#310
pLC-shuttle Ag-CEN5-ARS	#268
pHC-shuttle-Eco-**CEN1**	#C875
pHC-shuttle-Eco-**CEN2**	#C876
pHC-shuttle-Eco-**CEN3**	#C877
pHC-shuttle-Eco-**CEN4**	#C878
pHC-shuttle-Eco-**CEN7**	#C879
pHC-shuttle-Eco-**CEN8**	#C880

FIG. 5.—In vivo assay for centromere activity. *Holleya sinecauda* was transformed with ARS-plasmids additionally containing regions harboring *Eremothecium coryli* centromere loci as indicated. Control plasmids with only an ARS give rise to small and irregular colonies. The addition of centromere DNA (*AgCEN5*) leads to faithful plasmid segregation of plasmids and results in large colonies. Nonfunctional *E. coryli* centromere loci are marked by asterisks. Five initial transformants were repicked on selective plates and incubated at 30°C for 3 days prior to photography.

in *E. cymbalariae* and *A. gossypii*, but this gene has not been annotated in the yeast ancestor. The genes found linked in *E. coryli* are dispersed to two telomeres in *E. cymbalariae* indicating that this is a composite locus in *E. coryli*.

Eco	Ag	Ecym	Pre-WGD-ancestor chomosomes*			Ecym	Ag	Eco
S2R	6R	1L	L	Anc1	R	3R	5R	S1R
Internal S0	2L	6R	L	Anc2	R	5R	–	S11R
S3R	7R	2L	L	Anc3	R	6L	–	Internal S5
Internal S1	–	8R	L	Anc4	R	7R	4L	S0L
S9L	7L	5R	L	Anc5	R	5R+4L	–	S6R
S15L	–	3L	L	Anc6	R	7L	–	Internal S7
Internal S7	4R	5L	L	Anc7	R	2R	3L	S8R
S16R	2R	1R	L	Anc8	R	–	–	Internal S5
5	6	8	Sum of conserved ancestral telomeres			7	3	5

Fig. 6.—Identification of telomere loci in *Eremothecium coryli*. The positions of *Eremothecium* homologs of telomere linked genes of the pre-WGD ancestor were identified. In *E. cymbalariae* 15/16 ancestral telomere loci are conserved telomeres, for example, genes located at the left end of chromosome 1 (Anc_1L) of the yeast ancestor are found at *E. cymbalariae* chromosome 1L, and genes at Anc_1R are found at *E. cymbalariae* chromosome 3R. Genes from Anc_5R were relocated between two telomeres in *E. cymbalariae* (5R+4L). Lack of conservation of telomere positioning is indicated as (—). In *Ashbya gossypii*, 9/15 telomere loci are conserved. For analysis of *E. coryli*, the assembled scaffolds were used. Here, telomere linked genes were found at the end of 10 scaffolds. The remaining six ancestral telomere positions were found within scaffolds (e.g., intS5). Note two scaffolds (S5 and S7) were identified twice—directing our search for telomere-to-telomere fusion events in *E. coryli*.

In the yeast ancestor Anc_7.1 encodes a glutamate dehydrogenase, the *S. cerevisiae* ortholog of *YAL062W/GDH3*. This gene is absent from both *A. gossypii* and *E. cymbalariae*. Interestingly, this gene has been conserved in *E. coryli* at the junction of the telomere fusion. The gene is functional and conveys growth to *E. coryli* using ammonium sulfate as sole nitrogen source. Minimal media for growing *A. gossypii* or *E. cymbalariae* are supplemented instead with asparagine as nitrogen source as they cannot grow in standard minimal medium without amino acids and with ammonium sulfate generally used for *S. cerevisiae* propagation (to be published elsewhere).

Next to *E. coryli GDH3* two tRNAs are located. This suggests that the telomere-to-telomere fusion may have been brought about by homologous recombination involving these tRNAs rather than by head-to-head fusion of two telomeres (fig. 7B).

Assembly of the *E. coryli* Genome

The initial assembly of the *E. coryli* genome provided 19 scaffolds. Using conserved/ancient synteny, we aligned these scaffolds into six chromosomes. This required linking of scaffolds at 13 positions. In seven cases, these assignments were based on synteny with the other *Eremothecium* species and the pre-WGD ancestor. One other case was *Eremothecium* specific regarding the duplication of *FLO5* (*AFL092C/AFL095C*).

Another one involved synteny at the rDNA-repeat locus. The remaining four cases involved reciprocal translocations. For chromosome 6, two single reciprocal translocations can be inferred. One involved the *A. gossypii* homologs *AGL220W-AER272C* and *AGL219W-AER273C* whereas the other occurred between *AER168C-ABL066C* and *AER169C-ABL065W*. More than one reciprocal translocation is required to generate chromosome 1. In this case, both tRNA sequences and LTRs can be found at the scaffold ends, which generated difficult regions for automated assembly and regions that were also not covered by the 8 kb library used for sequencing. We conclude that based on the low number of scaffolds and by using comparative genomics, the assembly of the *E. coryli* genome into six chromosomes can be done (fig. 8). We thus assigned systematic names to all identified *E. coryli* genes based on their position in this assembly, for example, Eco_1.001 for the first ORF at the left end of chromosome 1 counting up to the right end of chromosome 1 harboring Eco_1.514 (see supplementary material, Supplementary Material online).

Based on this assembly, the *E. coryli* chromosomes are between 985 and 2,330 kb in size. We identified three mating type loci: A presumably active *MATα* and a telomeric *HMLα* on chromosome 2 and a telomeric *HMRa* on chromosome 4. The dispersal of mating type loci to different chromosomes has also been found in *A. gossypii*, whereas in *E. cymbalariae* all
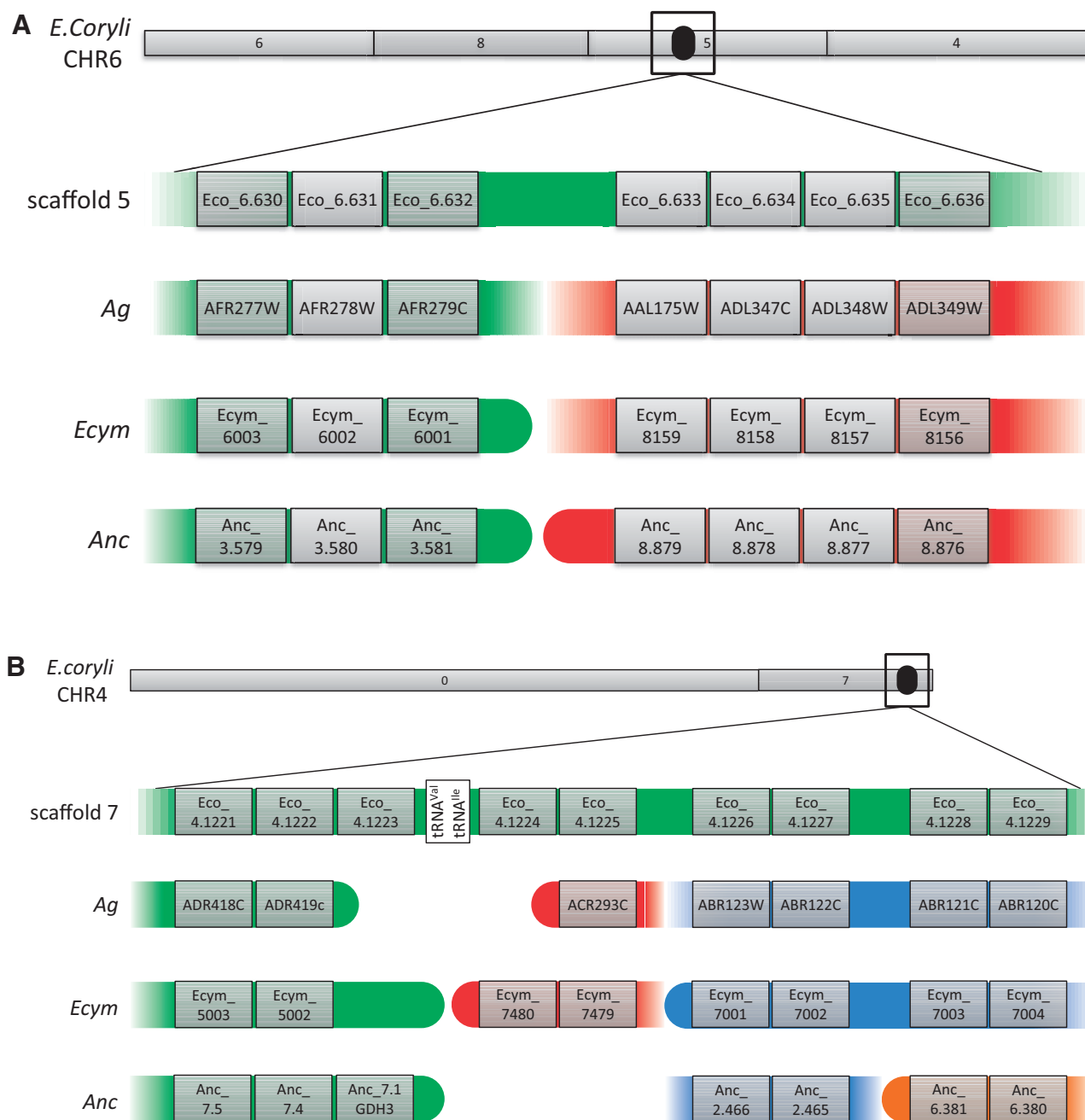
Fig. 7.—Telomere-to-telomere fusion events in *Eremothecium coryli*. Two loci indicative of telomere-to telomere fusion in *E. coryli* were identified on scaffolds 5 and 7. The order of *E. coryli* genes of scaffold 5 on CHR6 (*A*) and scaffold 7 on CHR4 (*B*) is shown aligned with homologs from *Ashbya gossypii*, *E. cymbalariae*, and the pre-WGD ancestor. Telomere ends are drawn with round-shaped edges, internal regions are depicted as open bars. Positions of *E. coryli* genes on the assembled *E. coryli* chromosomes are shown. Numbers within the *E. coryli* chromosomes correspond to the contributing scaffolds (see also fig. 8).

three mating type loci are located on chromosome 1 (Wendland and Walther 2005, 2011; Dietrich et al. 2013).

Assembly of an ERA

Eremothecium coryli now presents the third *Eremothecium* genome that has been sequenced next to completion. Due

to the large degree of synteny and with the ability to compare gene order with the reconstructed pre-WGD ancestor, we aimed at reconstructing individual segments of an ERA. We used a manual parsimony approach based on block synteny. We started at the eight centromere loci and assembled synteny blocks in both directions toward the telomeres. At
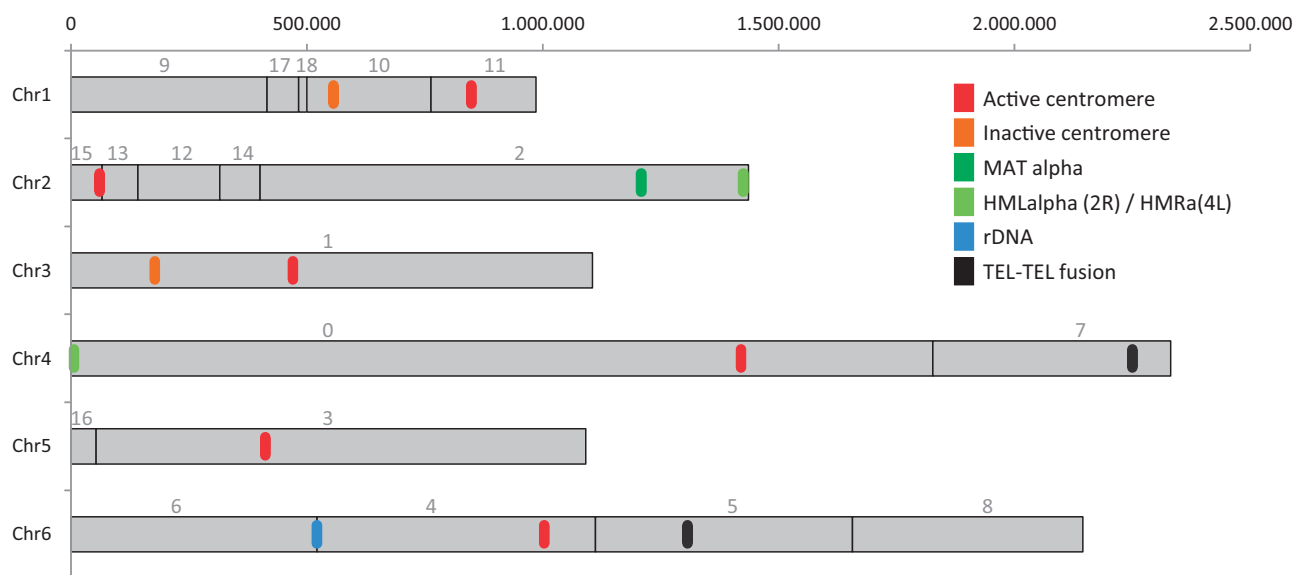
Fig. 8.—Assembly of *Eremothecium coryli* chromosomes. The 19 scaffolds from the original assembly left 13 gaps. Scaffolds were conceptually linked based on conserved synteny, which closed 10 gaps. Single reciprocal translocation closed one gap (between scaffolds 4 and 5). The remaining two gaps on chromosome 1 (scaffolds 11 and 10 and scaffolds 10 and 18) were linked by a set of reciprocal translocations. The size of each chromosome is according to scale. Scaffolds (also to scale) merged into chromosomes are indicated above the individual chromosomes. Key genome features are shown in the legend.

breakpoints of synteny in one *Eremothecium* species or the pre-WGD ancestor, the conserved gene order of at least two *Eremothecium* genome assemblies was relied on. This generated a telomere-to-telomere assembly of three ERA chromosomes, termed CHR3, CHR4, and CHR7 based on the founding centromeres (fig. 9). ERA_CHR3 contains 701 genes, ERA_CHR4 451 genes, and ERA_CHR7 732 genes in this assembly (see supplementary material, Supplementary Material online). At positions were all *Eremothecium* genomes differ among themselves and compared with the pre-WGD ancestor no conclusive progression could be called. Inclusion of further *Eremothecium* genomes will be required to improve this ERA assembly.

However, the ERA chromosome assembly of at present three chromosomes allows a view on the series of rearrangements that led from the ERA to the present-day *Eremothecium* species. Interestingly, this shows that the *E. coryli* genome is more syntenic to ERA than either of the other *Eremothecium* species or the pre-WGD ancestor, whereas *A. gossypii* harbors the most rearranged genome of these *Eremothecium* species (fig. 9).

Discussion

Once the yeast genome project was finished the wealth of information that can be drawn from a genome project became immediately clear (Goffeau et al. 1996). One striking result was the discovery of duplicated groups of genes on chromosome XIV and, more comprehensively, the WGD

(Philippsen et al. 1997; Wolfe and Shields 1997). The yeast genome sequence was instrumental in getting other genome sequencing efforts under way. Particularly the genomes of *A. gossypii* and *Lachancea waltii*, two protoploid, "pre-WGD," species, reinforced the concept of genome evolution by a WGD in the *Saccharomyces* lineage (Dietrich et al. 2004; Kellis et al. 2004). With an increasing number of complete genomes and draft genome sequences available for the *Saccharomyces* lineage, it became possible to reconstruct a yeast ancestral genome as it may have existed just prior to the WGD based on syntenic gene order conservation (Gordon et al. 2009).

The *Saccharomyces* complex has been resolved into 14 clades with clade 12 representing the genus *Eremothecium* (Kurtzman and Robnett 2003). This genus harbors both dimorphic (*E. coryli* and *H. sinecauda*) but also true filamentous fungi (*A. gossypii* and *E. cymbalariae*). The genus is of 2-fold commercial interest. *Ashbya gossypii* has long been known as an overproducer of riboflavin but species of this genus cause yeast spot disease or stigmatomycosis (Stahmann et al. 2000; Dietrich et al. 2013). For dispersal plant-feeding insect vectors of the suborder *Heteroptera* are used. A very persuasive hypothesis on how *Ashbya* developed into a riboflavin overproducer has been put forward: Some insects may be enabled to feed on toxic alkaloid-producing plants such as oleander when harboring *Ashbya*, whose riboflavin detoxifies these alkaloids and thus opens this ecological niche for both fungal and insect species (Dietrich et al. 2013).
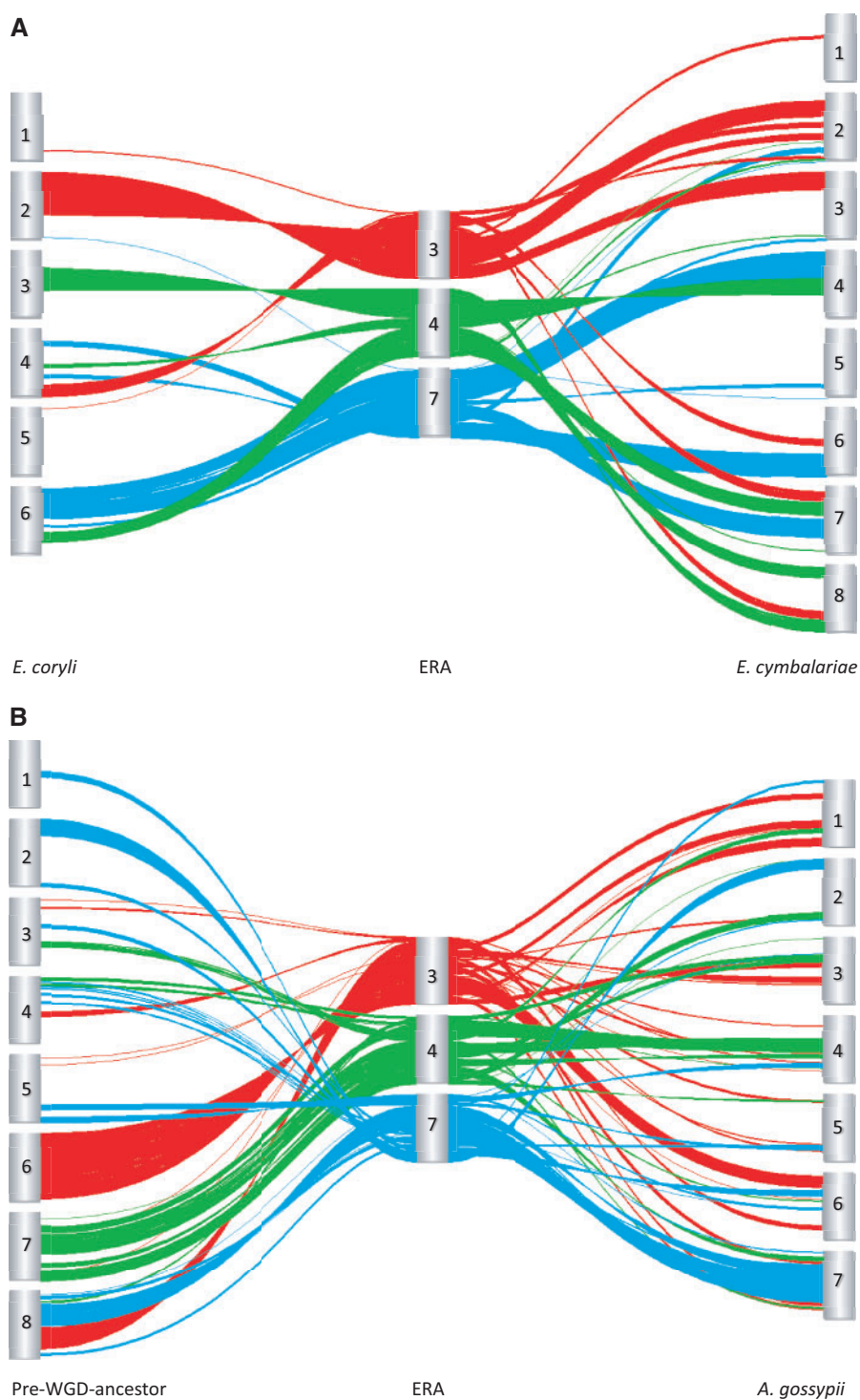
Fɪɢ. 9.—Comparative view of genome rearrangements. The compiled ERA was compared with the pre-WGD ancestor and *Ashbya gossypii* (*A*) and to *Eremothecium coryli* and *E. cymbalariae* (*B*). Each pair of homologous genes is linked by one line between the genomes—consecutive blocks of homology show as bars. The more individual lines emanating from ERA toward one genome the more genomic rearrangements occurred. This identifies *E. coryli* with the least number of rearrangements and *A. gossypii* with most rearrangements (for full details, see supplementary material, Supplementary Material online). Strudel software (http://bioinf.hutton.ac.uk/strudel/, last accessed May 15, 2014) was used to generate the overviews.

Here, we have sequenced the first dimorphic *Eremothecium* species. Based on synteny, we identified eight *E. coryli* loci homologous to *E. cymbalariae* centromere loci. Previously, the heterologous function of *A. gossypii* centromere DNA in *H. sinecauda* was shown (Schade et al. 2003). Using this assay, we could show that *CEN1* and *CEN8* were decommissioned in *E. coryli*. Concomitantly, we identified two sites of telomere-to-telomere fusion based on conserved sequences located to telomeres in *E. cymbalariae* and the pre-WGD ancestor (Gordon et al. 2011; Wendland and Walther 2011). Interestingly, *CEN8* in *A. gossypii* has also been eliminated. However, the mechanism has been different. Instead of a telomere-to-telomere fusion in *Ashbya* a break (or nonreciprocal translocation) at the centromere and fusion of the two chromosome arms to two different telomeres occurred. The consequences of this restructuring of *CEN8* are unclear. Yet, since *E. coryli* is a dimorphic fungus (lacking the characteristic Y-shaped dichotomous tip branching) and *A. gossypii* is a true filamentous fungus, we do not consider these events to be decisive for the evolution of hyphal growth—also given that the filamentous *E. cymbalariae* possesses a functional *CEN8*.

Eremothecium CEN8 has been assigned to chromosome 5 of the pre-WGD ancestor (*Anc_CEN5*), whereas *CEN1* of *Eremothecium* corresponds to *Anc_CEN1*. *Anc_CEN5* was also lost in *Candida glabrata*. Similarly, *Anc_CEN1* was lost in *C. glabrata* and also in *Vanderwaltozyma polyspora* (Gordon et al. 2011).

The internalization of telomeres, for example, via telomere-to-telomere fusions may preserve genes by placing them in a genomic context that may constrain their further evolution or alteration of expression patterns compared with more rapidly evolving telomeric loci (Teixeira and Gilson 2005; Batada and Hurst 2007; Ottaviani et al. 2008). In the case of the Anc6R-Anc7L fusion in *E. coryli*, a homolog of glutamate dehydrogenase (*ScGDH3*) was retained that has been lost in *A. gossypii* and *E. cymbalariae*. *EcoGDH3* enables *E. coryli* growth in media containing ammonium sulfate as sole nitrogen source. Similarly, via internalization of telomere Anc4L in *E. coryli*, a homolog of a *Lachancea thermotolerans* gene with similarity to a zinc-finger transcription factor (*ScRDS1*) has been retained.

With the currently available genome sequences of *Eremothecium* species and in combination with the pre-WGD ancestor, the reconstruction of an ERA was initiated and generated three of the eight chromosomes. This ancestral karyotype allows insight into chromosomal evolution that occurred within the *Eremothecium* lineage and also in comparison to other genera of the *Saccharomyces* complex. The *E. coryli* genome is more syntenic to ERA than the filamentous *Eremothecium* species. This may suggest that the ERA was a unicellular/dimorphic yeast whereas true hyphal growth is an apomorphy in the *Eremothecium* lineage. The independent evolution of hyphal growth in different ascomycetous lineages will fuel future comparative mechanistic studies to understand the molecular wiring of hyphal growth.

Paleogenomic studies of reconstructing ancestral karyotypes may provide hints of decisive evolutionary steps in a lineage (Yegorov and Good 2012). Comparison of lineage-specific ancestral genomes may provide insight into evolutionary steps at branch-points in phylogenetic trees. This directs future research to positions of synteny breaks, for example, between ERA and the pre-WGD ancestor for gene functions or changes in gene regulation that may have distinguished the *Eremothecium* clade from other *Saccharomycetes* in terms of filamentous growth, sporulation, or general metabolism.

Finally, by using build-a-genome methodologies, it has been demonstrated that synthetic DNA segments can be assembled (Dymond et al. 2009, 2011). With this technology even complete synthetic ancestral genomes could be generated and studied in the future.

Supplementary Material

Literature Cited

Ashby SF, Nowell W. 1926. The fungi of stigmatomycosis. Ann Bot. 40: 69–83.

Batada NN, Hurst LD. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. Nat Genet. 39: 945–949.

Bhutkar A, Gelbart WM, Smith TF. 2007. Inferring genome-scale rearrangement phylogeny and ancestral gene order: a *Drosophila* case study. Genome Biol. 8:R236.

Borzi A. 1888. *Eremothecium cymbalariae*, nuovo ascomicete. Bull Soc Bot Ital. 20:452–456.

Dietrich FS, et al. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science 304: 304–307.

Dietrich FS, Voegeli S, Kuo S, Philippsen P. 2013. Genomes of *Ashbya* fungi isolated from insects reveal four mating-type loci, numerous translocations, lack of transposons, and distinct gene duplications. G3 (Bethesda) 3:1225–1239.

Dymond JS, et al. 2009. Teaching synthetic biology, bioinformatics and engineering to undergraduates: the interdisciplinary Build-a-Genome course. Genetics 181:13–21.

Dymond JS, et al. 2011. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature 477:471–476.

El-Mabrouk N, Sankoff D. 2012. Analysis of gene order evolution beyond single-copy genes. Methods Mol Biol. 855:397–429.

Gastmann S, Dunkler A, Walther A, Klein K, Wendland J. 2007. A molecular toolbox for manipulating *Eremothecium coryli*. Microbiol Res. 162: 299–307.

Goffeau A, et al. 1996. Life with 6000 genes. Science 274(5287):546, 563–567.

Gordon JL, Byrne KP, Wolfe KH. 2009. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. PLoS Genet. 5: e1000485.

Gordon JL, Byrne KP, Wolfe KH. 2011. Mechanisms of chromosome number evolution in yeast. PLoS Genet. 7:e1002190.

Hansen EC. 1904. Grundlagen zur Systematik der Saccharomyceten. Zbl Bakt II Natur. 12:529–538.

Kato T, Park EY. 2012. Riboflavin production by *Ashbya gossypii*. Biotechnol Lett. 34:611–618.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428:617–624.

Köhli M, Buck S, Schmitz HP. 2008. The function of two closely related Rho proteins is determined by an atypical switch I region. J Cell Sci. 121: 1065–1075.

Kurtzman CP. 1995. Relationships among the genera *Ashbya, Eremothecium, Holleya* and *Nematospora* determined from rDNA sequence divergence. J Ind Microbiol. 14:523–530.

Kurtzman CP, Robnett CJ. 2003. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. FEMS Yeast Res. 3:417–432.

Miyao GM, Davis RM, Phaff HJ. 2000. Outbreak of *Eremothecium coryli* fruit rot in California. Plant Dis. 84:594.

Muffato M, Roest Crollius H. 2008. Paleogenomics in vertebrates, or the recovery of lost genomes from the mist of time. Bioessays 30: 122–134.

Murray AW, Szostak JW. 1983. Pedigree analysis of plasmid segregation in yeast. Cell 34:961–970.

Osawa H, Stacey G, Gassmann W. 2006. *ScOPT1* and *AtOPT4* function as proton-coupled oligopeptide transporters with broad but distinct substrate specificities. Biochem J. 393:267–275.

Ottaviani A, Gilson E, Magdinier F. 2008. Telomeric position effect: from the yeast paradigm to human pathologies? Biochimie 90:93–107.

Philippsen P, et al. 1997. The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIV and its evolutionary implications. Nature 387: 93–98.

Prillinger H, et al. 1997. Phytopathogenic filamentous (Ashbya, Eremothecium) and dimorphic fungi (Holleya, Nematospora) with needle-shaped ascospores as new members within the Saccharomycetaceae. Yeast 13(10):945–960.

Schade D, Walther A, Wendland J. 2003. The development of a transformation system for the dimorphic plant pathogen *Holleya sinecauda* based on *Ashbya gossypii* DNA elements. Fungal Genet Biol. 40: 65–71.

Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33:W686–W689.

Schmitz HP, Philippsen P. 2011. Evolution of multinucleated *Ashbya gossypii* hyphae from a budding yeast-like ancestor. Fungal Biol. 115: 557–568.

Stahmann KP, Revuelta JL, Seulberger H. 2000. Three biotechnical processes using *Ashbya gossypii, Candida famata*, or *Bacillus subtilis* compete with chemical riboflavin production. Appl Microbiol Biotechnol. 53:509–516.

Steiner S, Wendland J, Wright MC, Philippsen P. 1995. Homologous recombination as the main mechanism for DNA integration and cause of rearrangements in the filamentous ascomycete *Ashbya gossypii*. Genetics 140:973–987.

Teixeira MT, Gilson E. 2005. Telomere maintenance, function and evolution: the yeast paradigm. Chromosome Res. 13:535–548.

Walther A, Wendland J. 2005. Initial molecular characterization of a novel Rho-type GTPase RhoH in the filamentous ascomycete *Ashbya gossypii*. Curr Genet. 48:247–255.

Wendland J, Walther A. 2005. *Ashbya gossypii*: a model for fungal developmental biology. Nat Rev Microbiol. 3:421–429.

Wendland J, Walther A. 2011. Genome evolution in the *Eremothecium* clade of the *Saccharomyces* complex revealed by comparative genomics. G3 (Bethesda) 1:539–548.

Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713.

Wünschmann J, et al. 2007. Phytochelatins are synthesized by two vacuolar serine carboxypeptidases in *Saccharomyces cerevisiae*. FEBS Lett. 581:1681–1687.

Yegorov S, Good S. 2012. Using paleogenomics to study the evolution of gene families: origin and duplication history of the relaxin family hormones and their receptors. PLoS One 7:e32923.

Associate editor: José Pereira-Leal