## Table S1 Sequencing metrics

The metrics of the PacBio and Illumina sequencing reads. 1) PacBio. Metrics of PacBio reads. Accession Number: Reads for assembly: BG2 SRR10877222; BG3993-1: SRR11102020. Resequencing reads: BG3993-2: SRR11788818 and SRR11788818. 2) Illumina. Metrics of Illumina genome sequencing reads for strain BG2. Accession Number: SRR10877325.

## Table S2 SNV and structural variants

1) SNV CBS138 BG2 BG3993. Metrics of SNVs and Indels between strain CBS138, BG2 and BG3993 determined using MUMMER3 (details in Methods). We further counted SNVs and Indels in coding regions, and variants resulting in nonsense mutations or loss of stop codons. Finally, we used BLASTP to count the amino acid changes. 2) Structural Variant Summary for strains CBS138, BG2, BG3993. Summary of structural variation (length difference > 50 nt) in strain CBS138 and BG3993 relative to strain BG2. 3) Structural Variation for strains CBS138, BG2, BG3993. Raw data of structural variant analysis using Assemblytics [(Nattestad and Schatz 2016)](http://sciwheel.com/work/citation?ids=3250709&pre=&suf=&sa=0) in strain CBS138 and BG3993 relative to strain BG2. 4) SNV density comparison. We compared the distribution of SNV density between different groups of regions using Kolmogorov–Smirnov statistics, and documented the p-values in this spreadsheet (details in Methods).

## Table S3 Chromosome rearrangement events

The rearrangement events are identified by analysis of non-syntenic ORFs between strains. The breakpoint of rearrangements is identified using BLASTN to compare the associated intergenic region in the rearranged strain to the reference strain. 1) BG2 vs CBS138. Identified rearrangement events in strain BG2 relative to the type strain CBS138. 2) BG3993 vs CSB138. Identified rearrangement events in strain BG3993 relative to the type strain CBS138, the structure shared between strain BG2 and BG3993, with the same rearrangement event relative to strain CBS138 is also documented. 3) BG3993 vs BG2. Identified rearrangement events in strain BG3993 relative to the strain BG2, the structure shared between strain BG2 and CSB138, with the same rearrangement event relative to strain BG2 is also documented.

## Table S4 Mobility of LTR

We identified the location of the LTRs associated with *TKP5* and *Ty3* in strain CBS138, BG2, BG3993, and documented the mobility of the LTRs by comparison of the LTR locations. In addition, we also analyzed the occurrence of LTRs in the DSY562 strain, sequenced by Vale-Silva *et al.* [(Vale-Silva *et al*. 2017)](http://sciwheel.com/work/citation?ids=7653877&pre=&suf=&sa=0). Detailed identification and comparison procedure in Materials and Methods. 1) TKP5 BG2, BG3993, CBS138. The locations of *TKP5* associated LTRs in strain BG2, CBS138 and BG3993. PIDE: percent sequence identity. The LTRs whose genomic location is shared with other strains are marked with “X”; self comparisons within a strain are masked in grey. We documented the loss of ancient LTRs (sequence identity < 95%) between strains. 2) Ty3 BG2, BG3993, CBS138. The locations of *Ty3* associated LTRs in strain BG2, CBS138 and BG3993. PIDE: percent sequence identity. The LTRs shared with other strains are marked with “X”, and the shared status for the same strain is masked in grey.

## Table S5 Subtelomeric structure

The subtelomeric structure of the 13 subtelomeres with different structures in BG2, BG3993 and CBS138. Large-scale rearrangement events may exchange the entire subtelomeres between chromosomes. Therefore, we renamed the subtelomeres to the subtelomeres in type strain CBS138 based on synteny information, *i.e.* the subtelomere name may not reflect the exact chromosome location in strain BG2 and BG3993. 1) Consensus Subtelomere. The ORFs in the putative consensus subtelomeres are based on BG2, BG3993, CBS138 (details in Materials and Methods). Source Strain: Source strain to identify the consensus ORF, *i.e.* strains in which the ORFs are located in the consensus subtelomere. We also referenced the DSY562 to identify consensus ORFs in a fourth strain. 2) BG2 BG3993 CBS138. The relative location of the subtelomeres. The 0 position for each subtelomere is the end of the last non-telomeric ORF. The names of the subtelomeres in BG2 and BG3993 are according to the CBS138 for subtelomeres entirely translocated to different chromosome ends, therefore, the subtelomeric names may not reflect the exact chromosome location. Tel: location of telomeric repeats identified using the telomere seed sequence (GGGGTCTGGGTGCTG). 3) DSY562. We used the assembly from Vale-Silva *et al.* [(Vale-Silva et al. 2017)](http://sciwheel.com/work/citation?ids=7653877&pre=&suf=&sa=0). The subtelomeric ORFs were extracted based on the last non-subtelomeric ORFs. The exception is ChrL left because it starts with non-subtelomeric ORFs, and we documented the last two terminal ORFs to represent ChrL Left. We also exclude the ChrL and ChrM Right, because their assembly does not contain the terminal ChrL and ChrM Right regions which encode the *EPA14a/b* ORFs. We updated the systematic names in the type strain CBS138 [(Xu *et al*. 2020)](http://sciwheel.com/work/citation?ids=8248789&pre=&suf=&sa=0), and we updated the names in DSY562 assembly for comparison. In their annotation, some systematic names were assigned to multiple ORFs in the DSY562 sequence [(Vale-Silva et al. 2017)](http://sciwheel.com/work/citation?ids=7653877&pre=&suf=&sa=0). For example, CAGL0I11011g3 indicates that it is the third ORF assigned to the CAGL0I11011g gene. We corrected these assignments by extraction of the N-terminal regions of those ORFs, and generated a N-terminal phylogenetic tree to obtain the corresponding name in our assembly. We also used synteny information to assign the systematic names. We documented the comparison of the ORF location with the consensus subtelomere structure in the Stats column. Consensus: Same structure in the consensus subtelomere. Removal of annotation: ORFs removed in our annotation ( putative ORFs are small and dubious). Inversion: ChrC Left undergoes a large-scale inversion event, resulting in ORFs in the chromosome body to be subtelomeric. Fragmented ORF: Degraded ALPs or fragments within the ORFs identified because of early stop codons. Non-subtelomeric: The ChrL Left subtelomere translocated to ChrD Left, leaving ChrL Left to end in non-subtelomeric ORFs. Translocation (BG2 specific ORF): CAGL0F09295g is only encoded in BG2, therefore, it is a translocation event relative to strain BG2, not to the consensus structure. Consensus with intragenic recombination (shared with BG3993). CAGL0F00099g undergoes an intragenic recombination with CAGL0F00077g in strain BG3993. Consequently, it becomes the terminal ORF. in BG3993. This structure is shared with strain DSY562.

## Table S6 Break induced replication

In order to inform whether the translocation events of the terminal ALPs resulted from the break induced replication (BIR), we evaluated the homology between the associated intergenic regions. We first identified the corresponding template locus and target locus (where the break occurs) of a translocated locus. For duplication events where novel ORFs are generated in consequence, we found the close homologs of the novel ALPs using the phylogenetic tree as the template ORFs. The template locus is the template ORF locus, and the target locus is the subtelomere of the novel ORF. For translocation events, the template locus is the consensus locus of the translocated ORF, and the target locus is the locus of the translocated ORF in the translocated strain. BIR leads to replacement of DNA sequence with a homology overlap around the breakpoint. Therefore, for each translocated intergenic region, we looked for the corresponding homologous region in the template and target locus of the reference strain which have the consensus structure. Except for CAGL0H10648g, which we cannot find a template ORF, five of the six “novel” or strain-specific ORFs are consistent with BIR, and we identified the length of homology overlap for BIR for four of them. We successfully identified the homology overlap for all the translocated ORFs. To avoid misleading subtelomeric names due to translocation of the entire subtelomere, the entirely translocated subtelomeres in BG2 and BG3993 are renamed according to the CBS138 subtelomeres. If there are multiple reference strains for one BIR event, we preferred the strain with longer homology overlap. Note that all the lengths of homologous regions are maximum “projection” length of the reference sequence over the query sequence, which are calculated simply as the subtraction of the query start and end coordinates of the alignments between the query and the reference sequences. Therefore, small insertions and deletions are not counted in homology lengths. Please refer to the Alignment Data for exact alignment results. In addition,

1) Template of Strain Specific ORF. QueryStrain and QuerySubtel: the strain and subtelomere where the strain-specific ORF is located. RefStrain: reference strain for homology analysis. TargetSubtel, HomoTargetLocus: target subtelomere in the reference strain and the length of the region in target subtelomere in reference homologous to the intergenic region in the query strain. TemplateSubtel and HomoLengthTemplate: template subtelomere in reference and the length of the region in template subtelomere in reference homologous to the intergenic region in the query strain. 2) Alignment Data for Strain Specific ORF. Raw BLASTN alignment of the query intergenic to the relevant reference subtelomeres (filtered for e-value < 10-6). 3) Translocated ORF. The same columns are used with those in sheet Template of Strain Specific ORF. Our consensus structure is computed assuming minimal rearrangement events, which may be problematic because we only computed the structure with three strains. Therefore, we also examined translocation in the reverse direction, which is from “non-consensus” structure to “consensus” structure. The reverse direction is used to calculate the homology distance if it has much better homology overlap for recombination. 4) Alignment Data for translocated ORF. Raw BLASTN alignment of the query intergenic to the relevant reference subtelomeres (filtered for e-value < 10-6). 5) Synonymous codon variation. SNVs present in the third position of codons of strain-specific (duplicated) ORF relative to the parental ORF within the same strain, and present in the parental ORF relative to the parental ORF orthologs in other sequenced strains. To enumerate the third position codon variation, the N-terminal region of the associated ALPs were pairwise aligned using BLASTP (e-value < 10-6). We then used the BLASTP alignment to identify shared amino acids between strains and for conserved amino acids, compared the third position of the codon in genome sequence. “Duplicated” indicates the strain-specific ORF (i.e. the product of putative BIR). “Parental” indicates the putative template ORF in the same strain. “Ortholog” indicates the ortholog of the parental ORF in a different strain. In Figure S4, we illustrated only one possible ortholog based on the BIR evidence in the intergenic region; here for completeness, we document divergence from both orthologues. Query and Subject are the query and subject used for BLASTP. Percentage of shared third codon variation was calculated as number of Third Codon Variants / number of Shared Amino Acids. We excluded the BG3993/CAGL0J12067g and BG2/CAGL0J12067g comparison to calculate the average variation of parental and ortholog pairs because BG2/CAGL0J12067g has an apparent higher variation, suggesting it is not an actual orthologue. BIR\_No 3 and 4 are the two BIR events with low third codon variation which is consistent with a relatively recent duplication.

## Table S7 Differences in ORF complement

We documented the ORFs which are not encoded in one of strains BG2, BG3993 or CBS138. 1) Major ORF Difference. Major change in encoded ORFs. Complex ChrC region: a complex region in the chromosome body of ChrC. This region is repetitive. Each repeat within the region encodes one ALP in adhesin cluster VIII and two small ORFs. Difference in the copy number of the repeat leads to difference in copy number of the encoded ORFs. 2) Other ORF Differences. We note four classes of ORF differences between strains: 1) the copy number variations of copper-binding metallothioneins; 2) copy number variation of HXT genes; 3) difference in the number of the retrotransposon inversion events; 4) change in annotation, which includes removal of short dubious ORFs as well as merging of two adjacent ORFs because of the stop codon mutation in the preceding ORF.

## Table S8 Oligos for fosmid cloning and sequencing

Oligos used for fosmid cloning and sequencing.