Chart, scatter chart

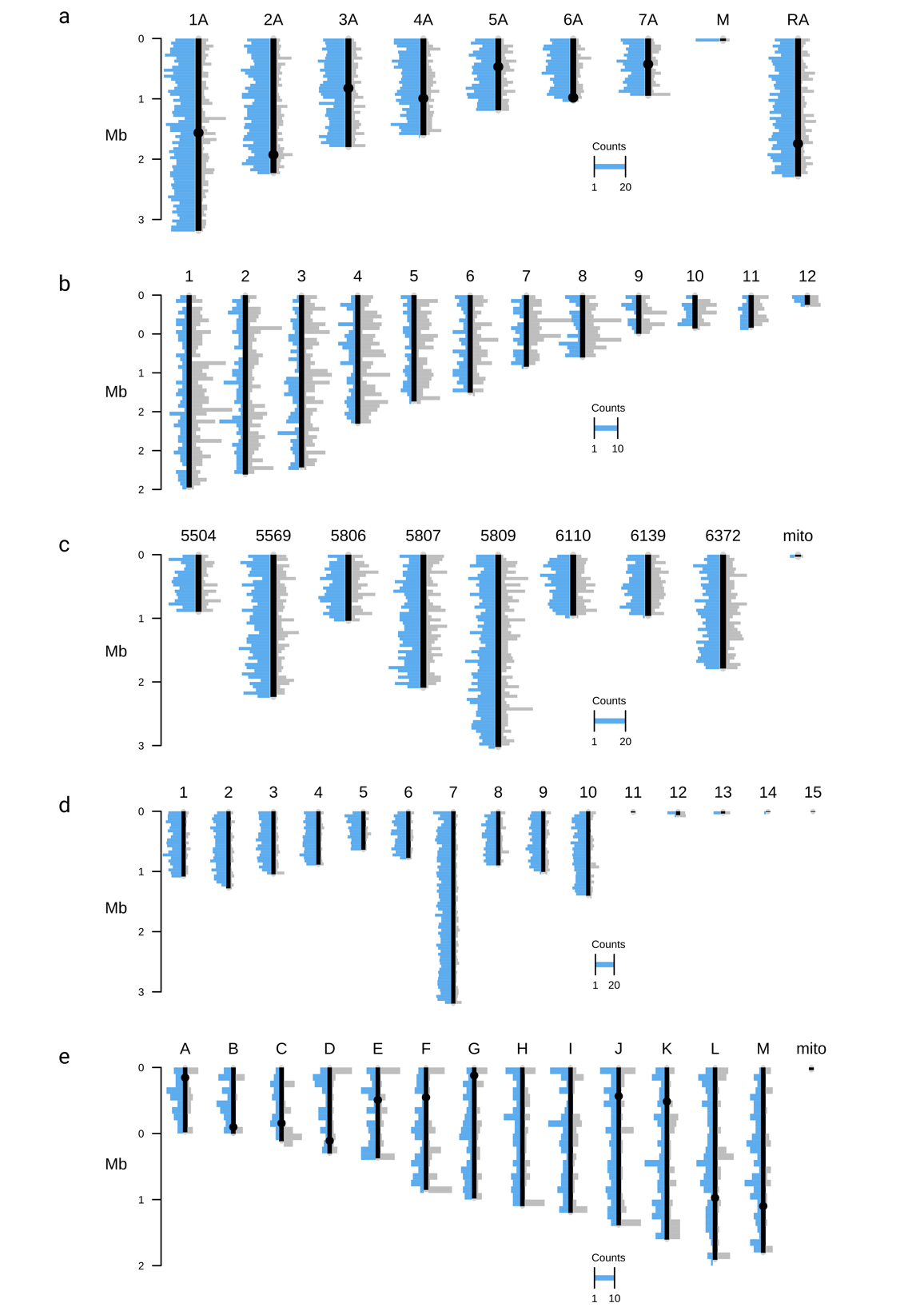
Description automatically generated

**Figure S1:** Principal Component Analysis (PCA) plots of all analyzed samples across species. Plots for **(a)** *C. albicans*; **(b)** *C. tropicalis*; **(c)** *C. parapsilosis*, **(d)** *C. auris* and **(e)** *C. glabrata*. Plots were generated using log2-transformed TPM expression values. Color codes correspond to SRA project accession numbers (or to sequencing read length for *C. auris*) and S dataset used in this study. Due to a large number of SRA projects and read lengths and for better visibility, the plot for *C. albicans* indicates only S and B datasets*.*

Chart, line chart

Description automatically generated

**Figure S2:** Saturation plots showing the number of identified lncRNAs depending on the number of analyzed samples. Per each species, samples were chosen randomly and the number of analyzed samples was incremented by 5 at each step.



**Figure S3:** Distribution of lncRNAs across chromosomes of the studied species. Plots for **(a)** *C. albicans*; **(b)** *C. tropicalis*; **(c)** *C. parapsilosis,* **(d)** *C. auris* and **(e)** *C. glabrata*. The chromosomes are depicted with black vertical lines, centromeres (for *C. albicans* and *C. glabrata*) are depicted with black circles. The number of antisense transcripts is depicted in blue (at the left side of chromosomes), of intergenic ones - in grey (at the right side of chromosomes). Each bar represents a histogram of number of lncRNAs across 50kb-long windows. For *C. tropicalis*, contigs smaller than 100 kb are not shown (i.e. contigs 13-23).

A picture containing scatter chart

Description automatically generated

**Figure S4:** For each Candida species, the number of both intergenic and antisense lncRNAs (represented by bar plots) relative to the distance to the closest telomere across 50 kb windows. The data is summed across all chromosomes. The red line across each plot represents the mean.

Chart, box and whisker chart

Description automatically generated

**Figure S5:** Comparison of different transcript properties of protein-coding genes (“pc”), lncRNAs found in the current study (“i” and “a” for intergenic and antisense transcripts, respectively), previously annotated lncRNAs (“known\_ncRNAs”), and intergenic regions (“inter”) in *C. glabrata.*

*Chart

Description automatically generated***Figure S6:** Percentage of lncRNAs overlapping repeat regions found in *Candida* genomes. Cumulative percentage for each species and lncRNA type can exceed 100% since one lncRNA can overlap more than one class of repeats. Numbers on the bar plot indicate the percentage (at the top) and number (at the bottom) of lncRNAs overlapping repeats.

*A picture containing diagram

Description automatically generated*

**Figure S7:** Assessment of evolutionary relationships of intergenic lncRNAs across *Candida* species. Each Venn diagram represents the number of classified lncRNAs families (top row) and the number of corresponding lncRNA in those families (bottom raw) across the species. The two used methodologies to generate the relatedness data corresponding to the Venn diagrams are mentioned at the top of the plot, namely BLAST-based classification on the left and secondary structure-based classification at the right.

A picture containing chart

Description automatically generated

**Figure S8:** Gene co-expression networks represented as dendrograms produced with WGCNA and hclust function based on 1-Topology Overlap Matrix. Networks for (a) *C. albicans*; (b) *C. tropicalis*; (c) *C. parapsilosis*, (d) *C. auris* and (e) *C. glabrata*. Each dendrogram represents co-expressed modules (on the bottom) obtained by gene clustering (top row) and eigengene clustering (bottom row).

Chart

Description automatically generated

**Figure S9:** Co-expressed modules and distribution of lncRNAs in modules for studied *Candida* species. Modules for (a) *C. albicans;* (b) *C. tropicalis;* (c) *C. parapsilosis*; (d) *C. auris* and (e) *C. glabrata*. Each bar plot represents a module, which height corresponds to the number of genes (both protein-coding and intergenic lncRNAs) involved in the module. Numbers at the right of bars represent the number of lncRNAs in the module and proportion (in parentheses) of lncRNA over the total number of genes in the module. Modules where a lncRNA represents a hub is highlighted by a red star.

Text

Description automatically generated

**Figure S10:** Expression levels of infection-specific lncRNAs and co-expressed protein coding genes across analyzed sample in (a) module “darkred” of *C. albicans* (for visualization purposes only samples from S dataset are shown) and (b) module coral1 of *C. tropicalis*. Only co-expression of infection-specific lncRNAs with highest WGCNA weights are shown. Sample IDs correspond to SRA identifiers and internal samples names in case of S dataset.

Diagram, venn diagram

Description automatically generated

**Figure S11:** Venn diagram of syntenic families where “infection-specific” lncRNAs of the four *Candida* pathogens are involved in.