**Material and Methods**

**Gene models**

Gene models in the *S. microadriaticum* genome were identified using a reference transcriptome assembly and *ab initio* prediction based on selected and refined gene structure models that were annotated with PASA (Haas 2003). First, all transcripts of the reference transcriptome (n = 58,592 transcripts, representing 41,679 distinct gene loci) were mapped to the genome assembly using PASA, yielding 22,827 annotated full-length gene structure models. In order to account for putative non-canonical splice sites in the gene structure models (i.e., GA-AG), we adapted the source code of PASA. From the set of 22,827 gene structure models, we applied several stringency criteria to filter and obtain a *bona fide* set of gene structure models that could be used as a training gene set for *ab initio* gene prediction using AUGUSTUS (version 3.0.2) (Stanke 2003, 2006). The subsequently applied filter steps were: (1) removal of incomplete genes without start and stop codon; (2) removal of genes with less than 3 exons; (3) removal of genes with ambiguous 5’ or 3’ untranslated regions (UTRs); (4) removal of redundant protein sequences in protein clusters as indicated by BLASTP (E-value: 1e-10), only the longest one was retained in every protein cluster; (5) removal of genes with repeat sequences as indicated by BLASTN alignments to a repeat library generated by RepeatScout. The final set of *bona fide* gene structure models contained 2,957 gene models on which AUGUSTUS was trained in order to predict genes models in the genome assembly using the default training pipeline. To improve the prediction accuracy, we also generated “hints” as supplementary evidence of gene presence and location by mapping all transcripts of the reference transcriptome to the genome assembly using BLAT and the respective scripts provided in AUGUSTUS. For the final *ab initio* gene prediction, the source code of AUGUSTUS was also slightly modified to account for non-canonical exon-intron boundaries. Finally, the set of *ab initio* genomic gene models was further refined using PASA as described in Baumgarten et al. (2015).