

Re-evaluation of *pir*, *fam-a* and *fam-b* multigene families

For *pirs*, we obtained four long-form and eight short-form clades as in [26] (tree 10) albeit with lower bootstrap support, possibly due to our overly stringent automated trimming of the sequence alignment (see Methods). With a few exceptions, *P. vinckei* *pir* genes majorly populated two clades - L1 and S7 and a subclade S1g. These clades, previously shown to be *P. chabaudi*-dominant, hold equal or near equal proportions of *P. vinckei* *pirs* too. The only other *P. chabaudi*-dominant clade, L4, remains as a completely *P. chabaudi*-specific gene expansion. No *P. vinckei* species- or subspecies-specific clades are evident except for two subclades that could be inferred as *PvvCY*-specific expansions within L1 and S7 (marked i and ii in tree 10). Speciation of *P. vinckei* subspecies from their common ancestor seems to have been accompanied by gene gain in L1 and S7 clades and gene loss in S1g subclade. There is an almost linear increase of around 20 genes in *PvvCY*, *PvbDA* and *PvIDE* in clade L1 *pirs* and a near doubling of clade S7 *pirs* in *PvpCR*.

The *fam-a* and *fam-b* phylogenies (tree 3 and 4 respectively) show that previously identified ancestral lineages [25] are maintained in *P. vinckei* too. The addition of *P. vinckei* genes resolved the ancestral clade of internal *fam-a* genes in chromosome 13 further into several well-supported *vinckei* group clades and a *berghei* group clade (marked as A in tree 3). The 19 other *fam-a* clades and five *fam-b* clades consisting of positionally conserved orthologous genes are also conserved in *P. vinckei* (pan-RMP clades marked with * in tree 3 and 4). The *fam-a* family has expanded in the non-Katangan *P. vinckei* subspecies through independent events of gene duplication in their common ancestor giving rise to several non-Katangan clades (marked as B in tree 3). There is only a moderate *P. vinckei*-specific expansion in *fam-b* giving rise to three clades (marked as A in tree 4) that includes *PvvCY* genes too, pointing to gene

duplications in the *P. vinckei* common ancestor. In both the phylogenies, species and subspecies-specific gene duplication events within the *vinckei* group are rare but do occur (marked as i-iv in tree 3 and 4).

***fam-d* multigene family**

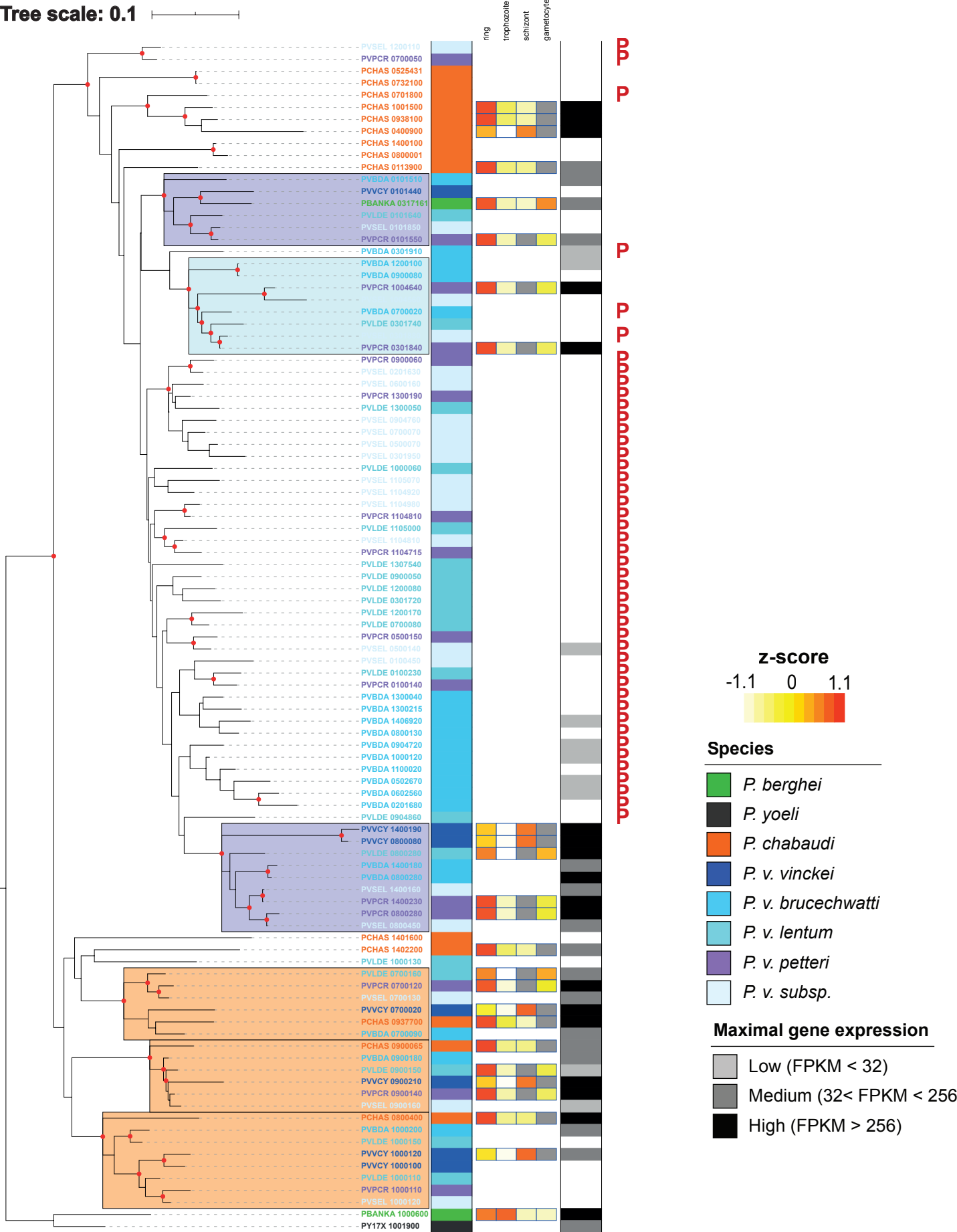
The *fam-d* multigene family is present as a single ancestral copy in *P. berghei* internally on chromosome IX but is expanded into a gene cluster in the same loci in *P. yoelii* (5 genes) and *P. chabaudi* (21 genes). Similar expansions have occurred in *P. vinckei* subspecies and phylogenetic analysis shows the presence of six robust clades within this family (tree 6). Clade I is clearly the ancestral clade from which all other *fam-d* genes have been derived as it consists of the single *P. berghei* gene and its orthologs in other RMPs, positionally conserved to be the outermost gene of the *fam-d* cluster in each RMP. While the *fam-d* family in *P. yoelii* is completely a product of paralogous expansion within Clade I, the *fam-d* families in the *vinckei* group seem to have expanded *via* five ancestral lineages forming clades II-VI. A subset of orthologs in Clade II (marked with * in tree 6) are positionally conserved among *vinckei* group parasites, located immediately after the *fam-d* ancestral copy and could therefore represent the Clade II ancestral gene in the *vinckei* group common ancestor. *PvvCY* has a smaller *fam-d* repertoire of 6 genes derived from only three of the five *vinckei* group lineages (Clade II, IV and VI), apart from the conserved ancestral copy.

Other multigene families

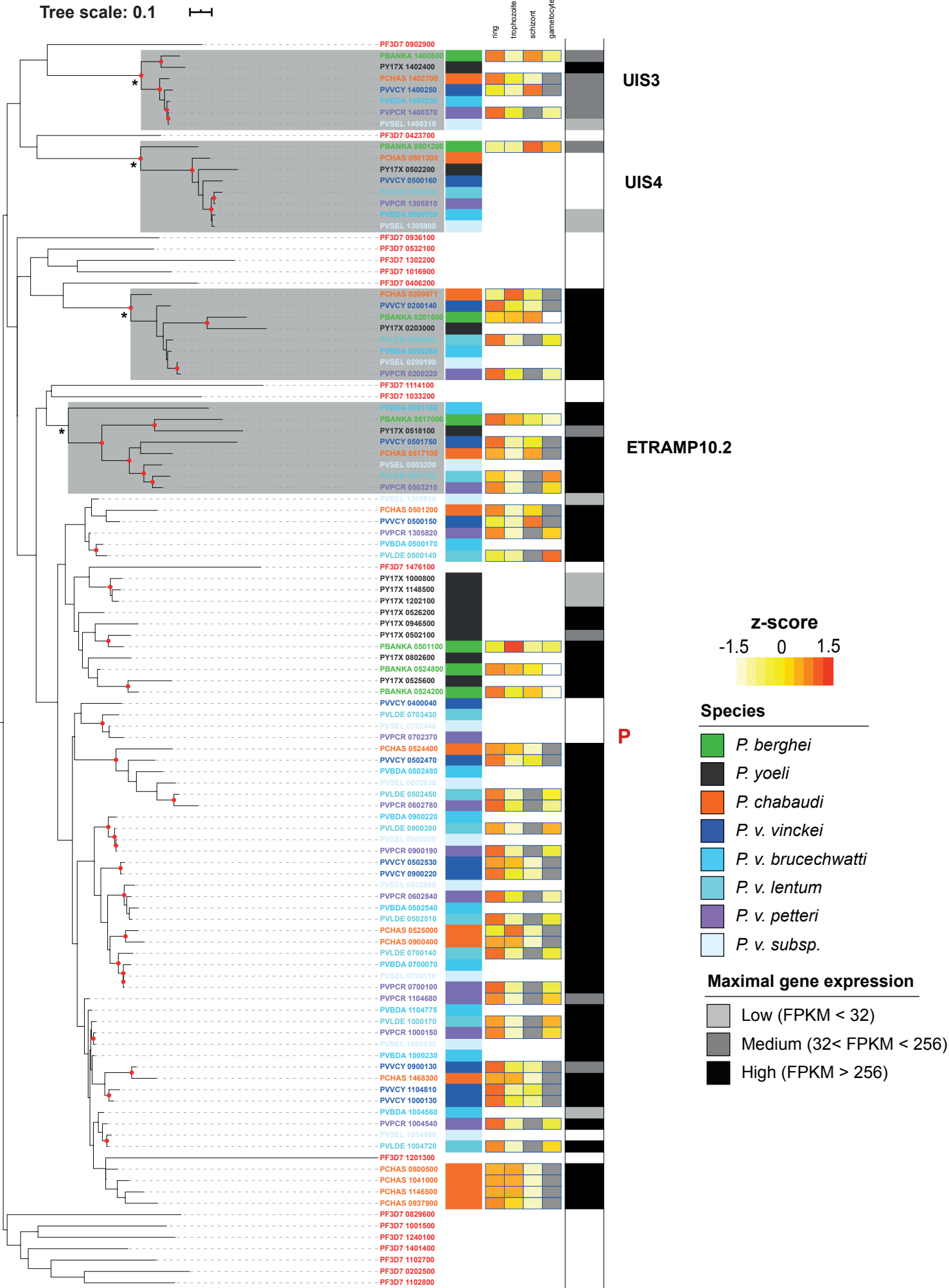
ML-based trees for haloacid dehalogenase-like hydrolase (*hdh*), putative reticulocyte binding proteins (*p235*) and lysophospholipases (*lpl*) have generally well-resolved topologies with robust bootstrap support for their nodes and some clades contain

syntenic orthologous genes (clades marked with * in trees 7, 8 and 9) to member genes in *P. falciparum*, for example, *PfHAD2*, *PfHAD3*, *PfHAD4* and *PfRH6*. Poor bootstrap support was obtained for the *etramp* tree (tree 2), however clades were identified for some members including *uis3*, *uis4* and *etramp10.2*.

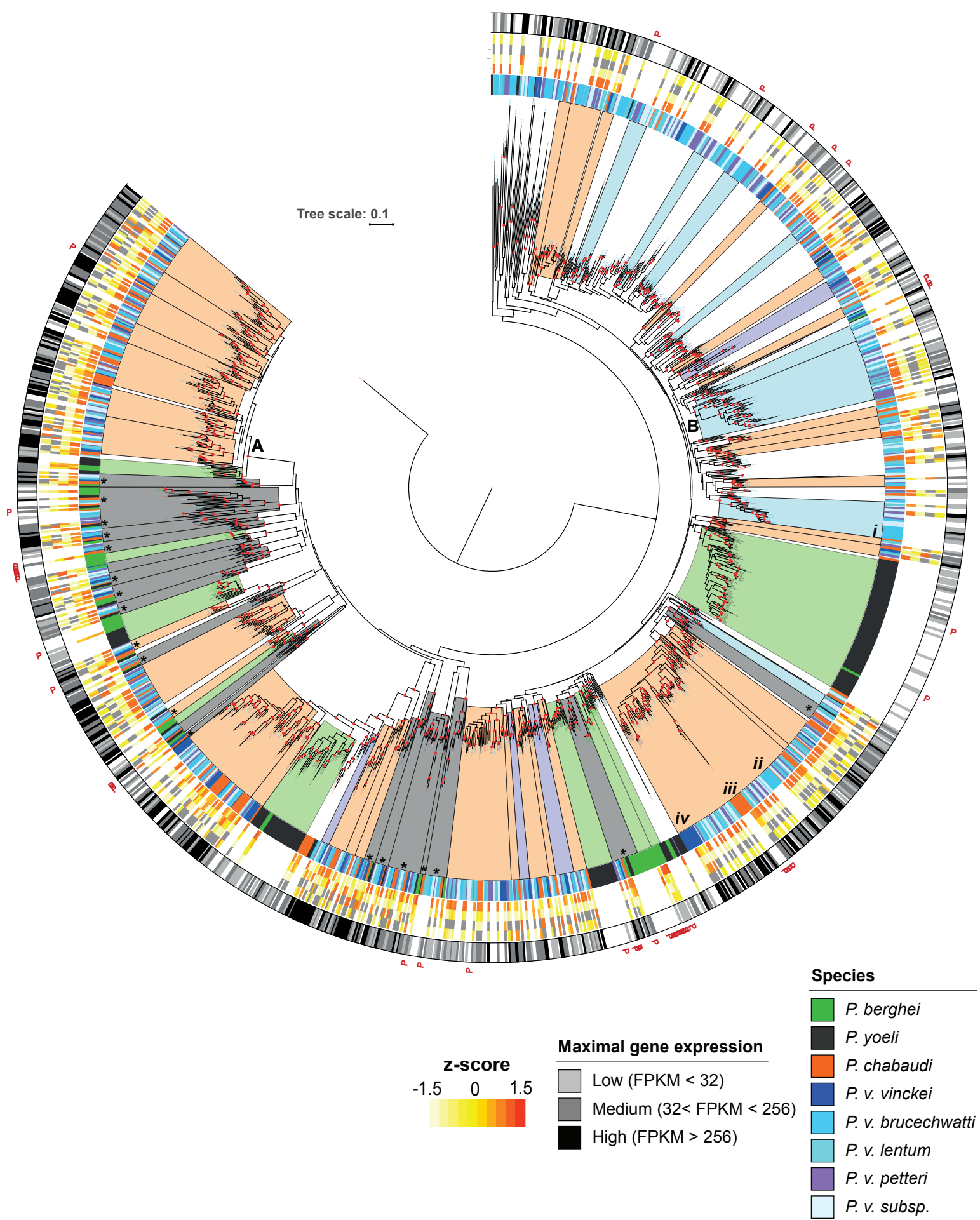
1. erythrocyte membrane antigen 1 (ema1)



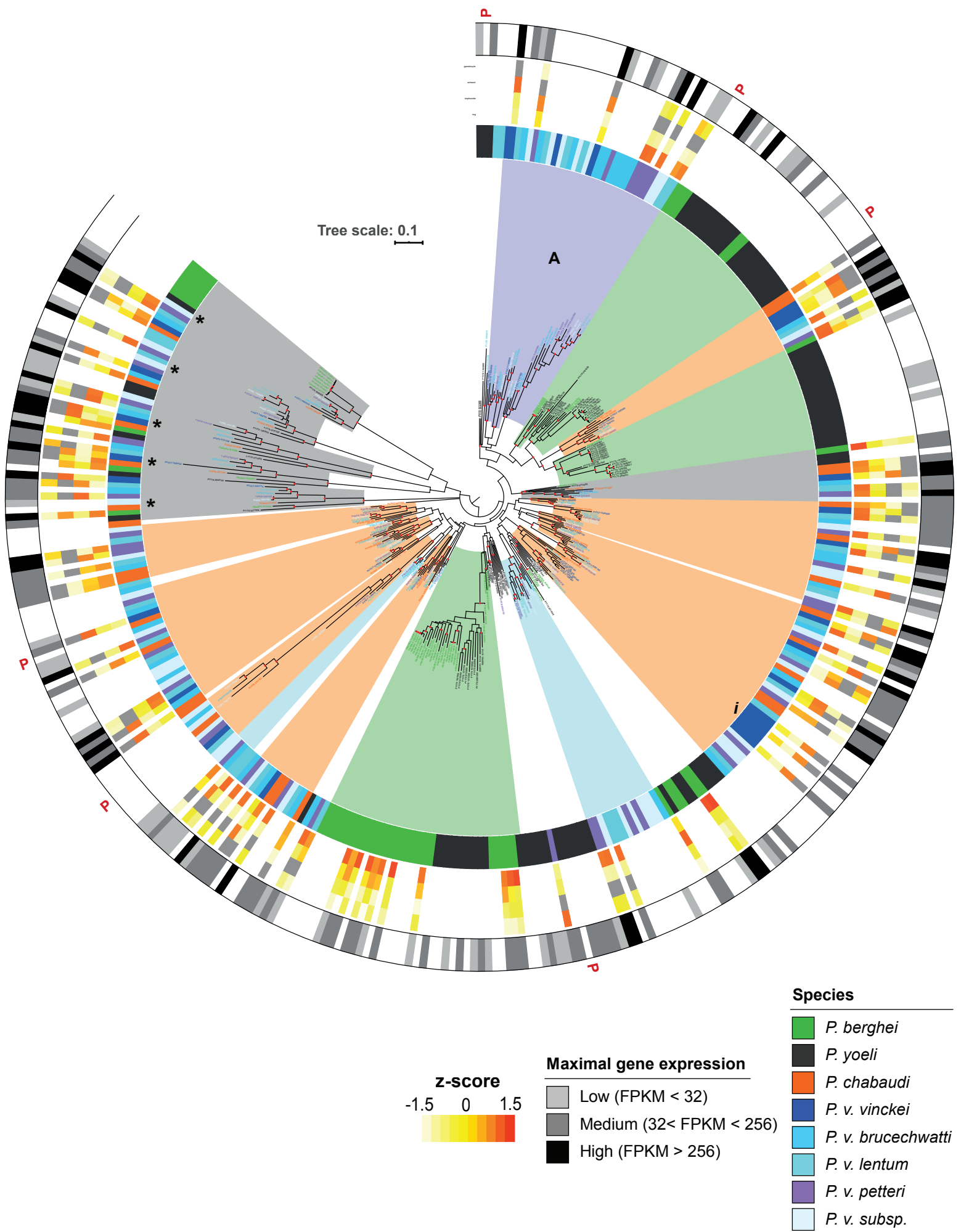
2. early transcribed membrane protein (*etramp*)



3. RMP-*fam-a*

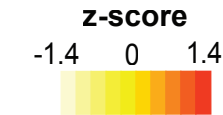
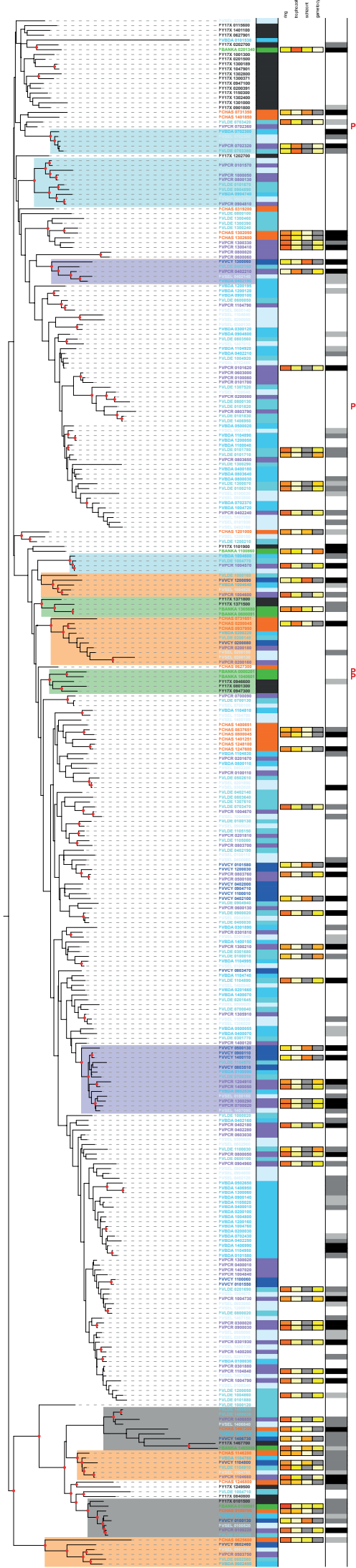


4. RMP-*fam-b*



5. RMP-*fam-c*

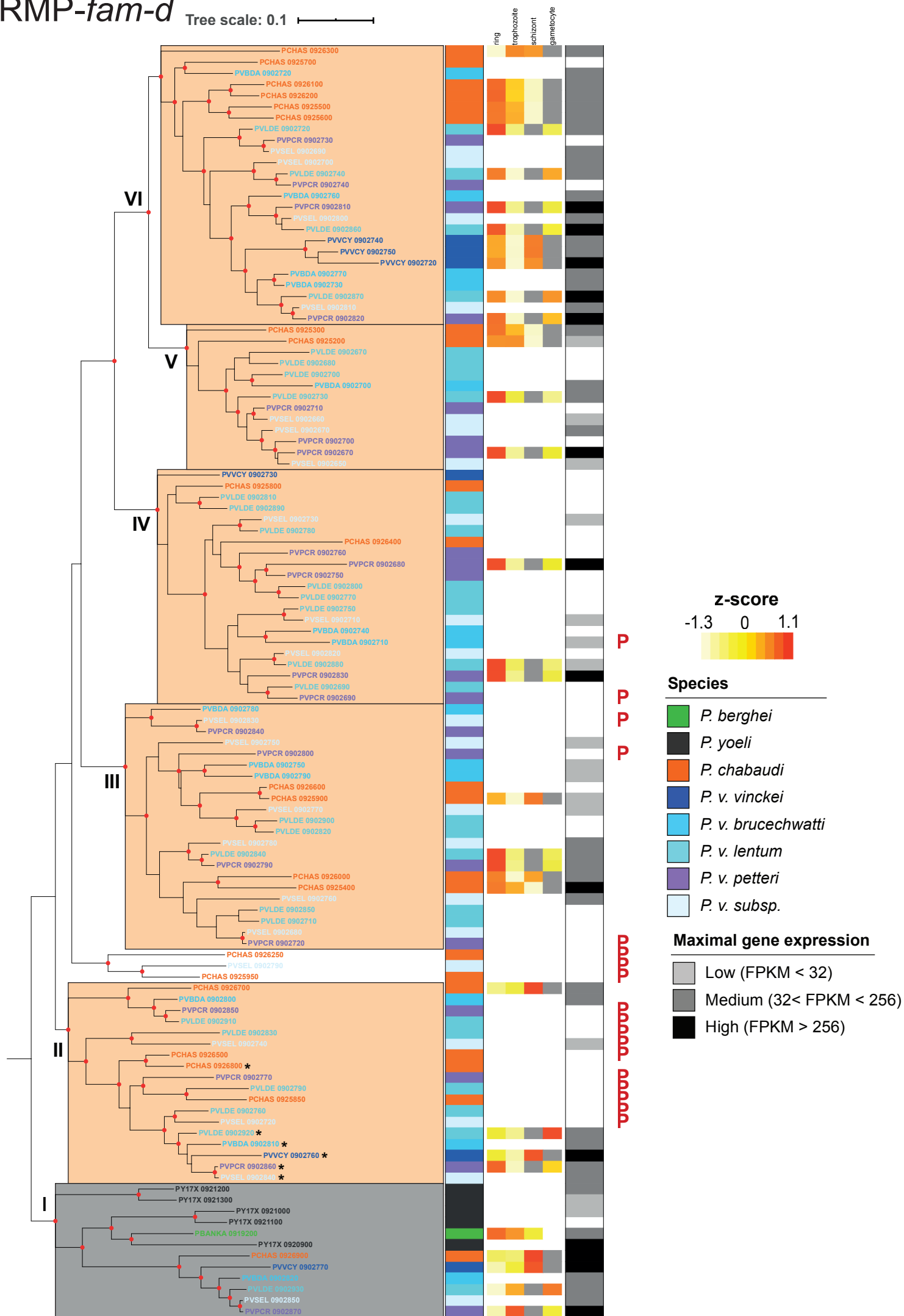
Tree scale: 0.1



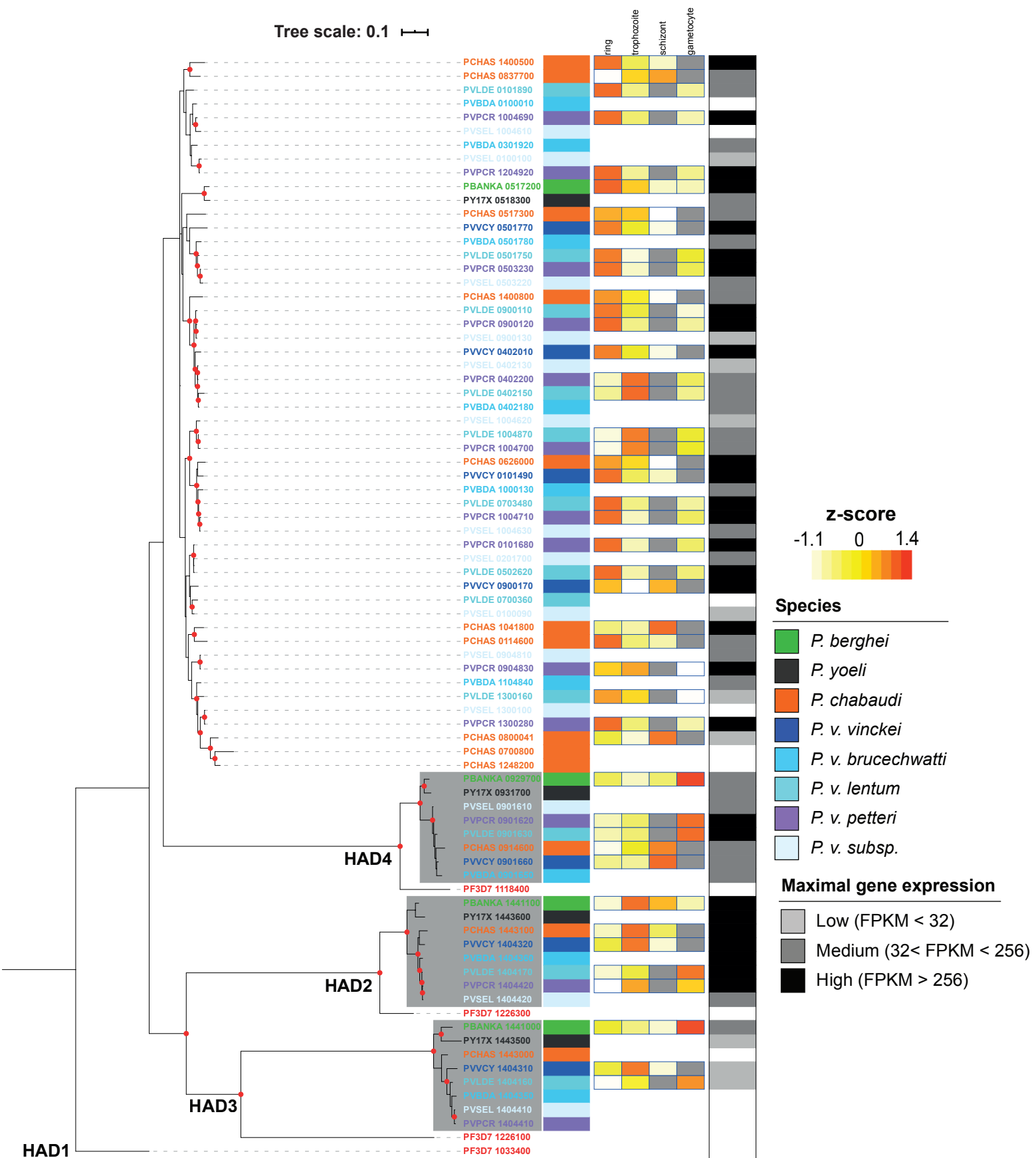
- Species**
- P. berghei*
 - P. yoeli*
 - P. chabaudi*
 - P. v. vinckei*
 - P. v. brucechwatti*
 - P. v. lentum*
 - P. v. petteri*
 - P. v. subsp.*

- Maximal gene expression**
- Low (FPKM < 32)
 - Medium (32 < FPKM < 256)
 - High (FPKM > 256)

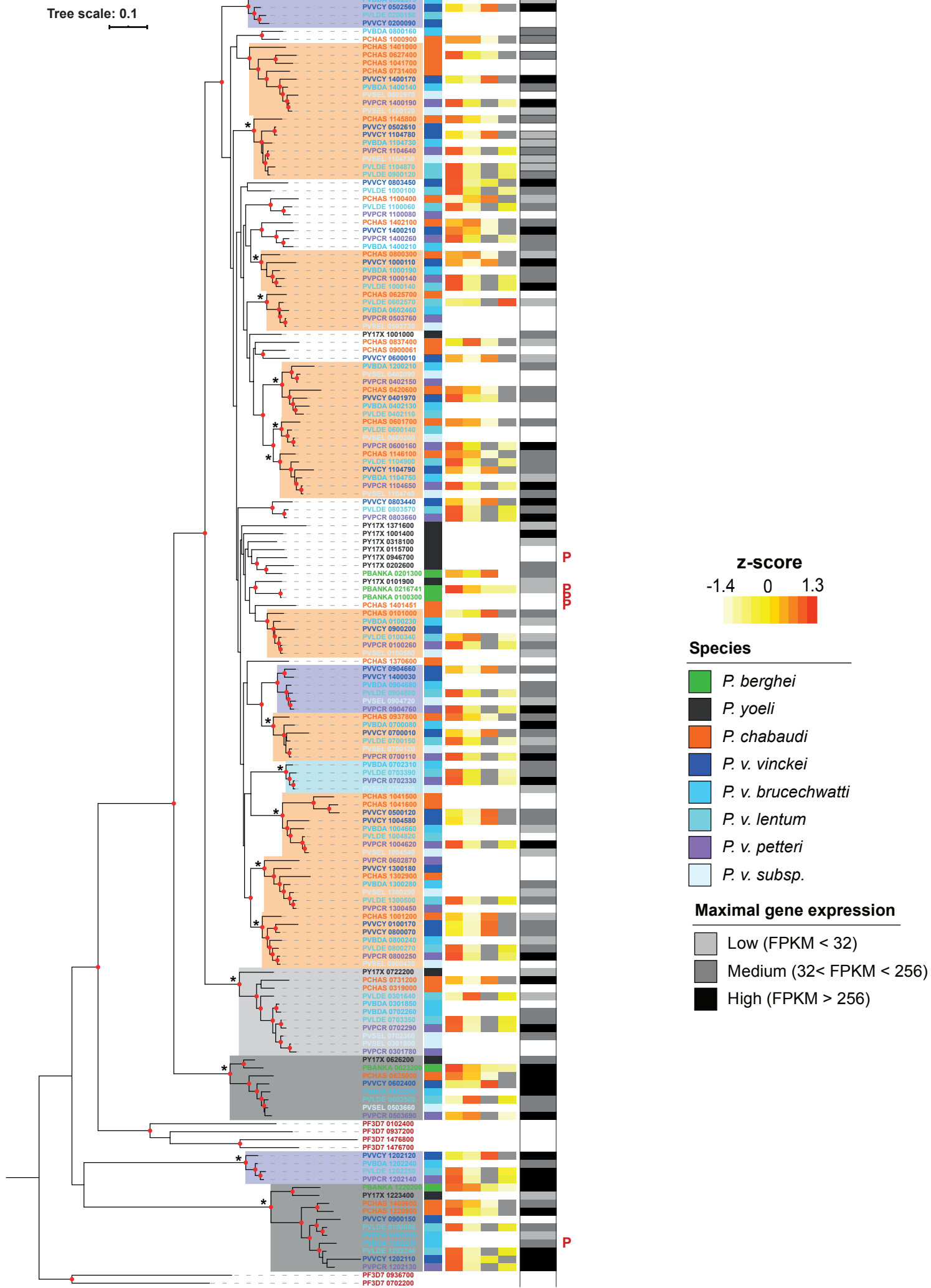
Tree scale: 0.1



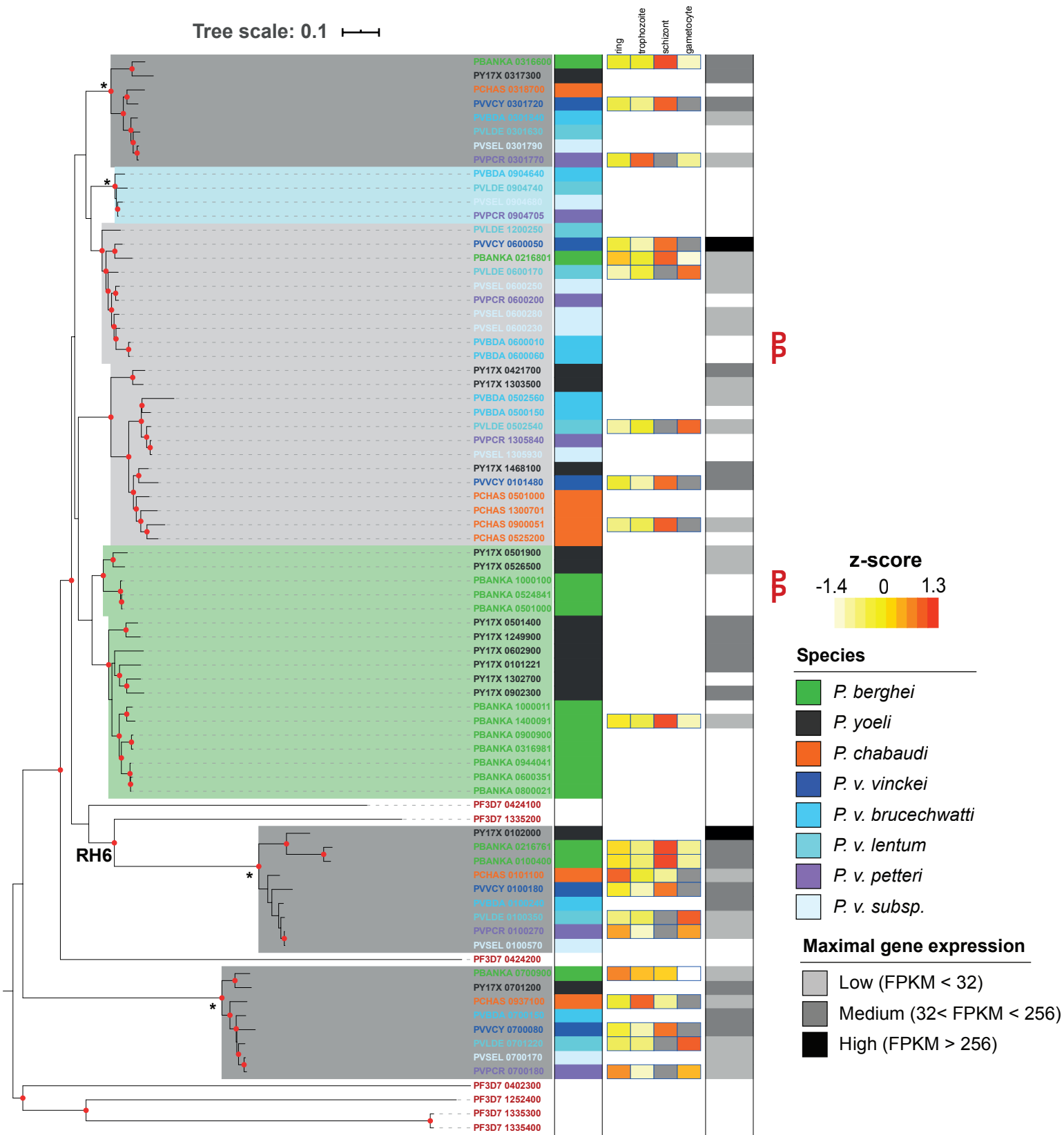
7. haloacid dehalogenase-like hydrolases (*hdh*)



8. lysophospholipases (*lpl*)



9. reticulocyte binding protein (p235)



10. *Plasmodium* interspersed repeat proteins (*pir*)

