

# Comparative Analysis of The Kinomes of *Plasmodium Falciparum*, *Plasmodium Vivax* And Their Host *Homo Sapiens*

Jack Adderley

RMIT University

Christian Doerig (✉ [christian.doerig@rmit.edu.au](mailto:christian.doerig@rmit.edu.au))

RMIT University

---

## Research Article

**Keywords:** Malaria, Plasmodium falciparum, Plasmodium vivax, kinome, kinase,

**Posted Date:** December 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1094331/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Genomics on March 26th, 2022. See the published version at <https://doi.org/10.1186/s12864-022-08457-0>.

# Abstract

**Background:** Novel antimalarials should be effective across all species of malaria parasites that infect humans, especially the two species that bear the most impact, *Plasmodium falciparum* and *Plasmodium vivax*. Protein kinases encoded by pathogens, as well as host kinases required for survival of intracellular pathogens, carry considerable potential as targets for antimalarial intervention<sup>1,2</sup>. To date, no comprehensive *P. vivax* kinome assembly has been conducted; and the *P. falciparum* kinome, first assembled in 2004, requires an update. The present study, aimed to fill these gaps, utilises a recently published structurally-validated multiple sequence alignment (MSA) of the human kinome<sup>3</sup>. This MSA is used as a scaffold to assist the alignment of all protein kinase sequences from *P. falciparum* and *P. vivax*, and (where possible) their assignment to specific kinase groups/families.

**Results:** We were able to assign six *P. falciparum* previously classified as OPK or 'orphans' (i.e. with no clear phylogenetic relation to any of the established ePK groups) to one of the aforementioned ePK groups. Direct phylogenetic comparison established that despite an overall high level of similarity between the *P. falciparum* and *P. vivax* kinomes, which will help in selecting targets for intervention, there are differences that may underlie the biological specificities of these species. Furthermore, we highlight a number of *Plasmodium* kinases that have a surprisingly high level of homology with their human counterparts and therefore not well suited as targets for drug discovery.

**Conclusions:** Direct comparison of the kinomes of *Homo sapiens*, *P. falciparum* and *P. vivax* sheds additional light on the previously documented divergence of many *P. falciparum* and *P. vivax* kinases from those of their human host. We provide the first direct kinome comparison between the phylogenetically distinct species of *P. falciparum* and *P. vivax*, illustrating the key similarities and differences which must be considered in the context of kinase-directed antimalarial drug discovery, and discuss the divergences and similarities between the human and *Plasmodium* kinomes to inform future searches for selective antimalarial intervention.

## Background

Despite significant progress in reducing malaria morbidity and mortality over recent decades, the disease remains a major health burden in tropical and subtropical regions, and puts significant strain on the medical and economic systems of heavily affected countries<sup>4</sup>. Malaria is caused by parasites of the genus *Plasmodium*, with *Plasmodium falciparum* and *Plasmodium vivax* being the most clinically relevant species. The World Health Organisation's current goal to achieve global elimination of malaria requires significant support for continued research and development into novel drugs with untapped targets to support the failing frontline treatments of the disease<sup>5</sup>. All currently deployed antimalarials are facing various levels of parasite resistance, which threatens to erode global progress towards eradication<sup>6</sup>. This is further marred by the lack of a highly efficacious vaccine (despite recent encouraging progress in this area) and by the ongoing SARS-CoV-2 pandemic that diverts public health resources away from malaria and other tropical diseases<sup>7-9</sup>.

This dire situation calls for the development of novel control agents with untapped modes of action that should be effective against both *P. falciparum* and *P. vivax*. One key class of signalling molecules, protein kinases, have been successfully targeted to treat numerous diseases; for example, 62 kinase inhibitors have reached the market for the treatment for a variety of conditions (prominently cancer), and this number keeps increasing<sup>10</sup>. Protein kinases encoded by pathogens, as well as host kinases required for survival of intracellular pathogens, carry considerable potential as targets for the treatment of infectious diseases such as malaria<sup>1,2</sup>. Protein kinases catalyse the transfer of phosphate from adenosine triphosphate (ATP) to a substrate polypeptide, resulting in structural and functional changes to the target protein. Phosphorylation events orchestrated by kinases in tandem with protein phosphatases play a pivotal role in signalling in all living cells. Many protein kinases are highly conserved even among evolutionarily distant organisms<sup>11</sup>. The kinome (i.e. the complement of all protein kinase-encoding genes) of *P. falciparum* was first assembled in 2004, reporting a total of 85 protein kinases, 65 of which, were related to typical eukaryotic protein kinases (ePKs)<sup>12</sup>. ePKs are distinguishable from atypical protein kinases (aPKs) through their conserved catalytic domain, which contains 12 distinct subdomains/motifs<sup>13</sup>. A majority of ePKs can be assigned to one of the following groups: CAMK, CMGC, AGC, TKL, TK, CK1 and RGC; ePK sequences that do not cluster in any of these groups by phylogenetic analysis comprise the “OPK” (other protein kinase) group; this includes important families such as the NEKs<sup>14</sup>. Over the past decades, the number of protein kinases identified in *P. falciparum* has steadily increased with the most recent study indicating a total of 105 protein kinases, (98 typical protein kinases and 7 atypical protein kinases, aPKs)<sup>1,15</sup>. Interestingly, the kinome of *Plasmodium vivax* has received considerably less attention, with, to our knowledge, only a brief overview available in the literature<sup>16</sup>, with no formally available phylogenetic tree of the parasite’s protein kinases or detailed comparison with the *P. falciparum* kinome; we intend to fill in this gap through the present study.

We utilised a recently published structurally-validated multiple sequence alignment (MSA) of the human kinome<sup>3</sup> as a scaffold to assist the alignment of all protein kinase sequences from *P. falciparum* and *P. vivax*, and (where possible) their assignment to specific kinase groups/families. Through this strategy we were able to assign six *P. falciparum* previously classified as OPK or ‘orphans’ (i.e. with no clear phylogenetic relation to any of the established ePK groups) to one of the aforementioned ePK groups (see below). The kinomes of *Homo sapiens*, *P. falciparum* and *P. vivax* were directly compared, shedding additional light on the previously documented divergence of many kinases, as expected in view of the phylogenetic distance between Opisthokonts and Alveolates (the clades that include metazoans and *Apicomplexa*, respectively), as well as new evidence for the conservation of a handful of kinases across both Opisthokonts and Alveolates. We also provide the first direct kinome comparison between the phylogenetically distant species of *P. falciparum* and *P. vivax*, illustrating the key similarities and differences which must be considered in the context of kinase-directed antimalarial drug discovery.

## Results And Discussion

**Protein kinase domain identification and assembly** -The *P. falciparum* sequences were collected from our earlier alignment <sup>1</sup>, while *P. vivax* sequences were obtained by searching the predicted *P. vivax* proteome in PlasmoDB v50b <sup>17</sup> using the term “kinase”. The resultant list was further refined to only include (i) sequences containing a Pfam ID of PF00069 or PF07714 (Protein kinase domain and protein tyrosine kinase, respectively) and (ii) sequences which contained the phrase “protein kinase” in ‘Product description’ data. These sequences were assessed using ScanProsite <sup>18</sup> to identify the protein kinase domain. Two regulatory subunits of CK2 (CK2 $\beta$ , both of which have characterised orthologues in *P. falciparum* <sup>19</sup>), a sequence annotated as “putative protein kinase” but that did not have a recognisable protein kinase domain, and a protein phosphatase, were removed (PVP01\_0904500, PVP01\_1212400, PVP01\_1030400 and PVP01\_1406400). Four atypical protein kinases (aPK) were identified: ABCK1 and ABCK2 from the ABC family (PVP01\_1430100, PVP01\_1334400), and Rio1 and Rio2, (PVP01\_1449100, PVP01\_0529500) all of which have orthologues in *P. falciparum* <sup>12</sup>. Finally, we identified a surprising four members of the phosphatidylinositol kinase (PIK) family in *P. vivax*, whereas this family is represented by only three enzymes in *P. falciparum*. It would be interesting to determine if this additional enzyme is implicated in vivax-specific biology, e.g. preference for reticulocyte or ability to establish dormant forms (hypnozoites) in hepatocytes. These sequences are PVP01\_1018600 (orthologous to PfPIK3), PVP01\_1024200 (orthologous to PfPIK4), PVP01\_0529300 (orthologous to PF3D7\_0419900) and PVP01\_1309200. A fifth *P. vivax* PIK-like kinase was initially identified (PVP01\_1404700), but its low Hidden Markov Model (HMM) score <50 (as defined by HMMER 3.3 <sup>20</sup>), and subsequent manual examination of the sequence, indicated this protein is highly divergent from the consensus PIK sequence and hence was not included here. Phylogenetic analysis of the PIKs kinase domains indicated that PVP01\_1309200 is divergent (**Supplementary Figure 1**). We further sought to determine if any additional protein kinases were encoded by *P. vivax* through a Psi-BLAST search (PlasmoDB’s Beta Blast interface – multiple parallel blast searches). We included the protein kinase domains of *P. vivax* (identified here), and those of *P. falciparum* <sup>1</sup> in the search query; however, no significant additional sequences were identified in this way. The typical protein kinase domains (78 in *P. vivax* [up from the 68 reported in the aforementioned preliminary study <sup>16</sup>], 98 in *P. falciparum*) were initially aligned using Clustal Omega <sup>21</sup> and imported into Jalview <sup>22</sup> for manual alignment adjustments.

**Hidden Markov Model (HMM) profiling for assignment of sequences to ePK families** – To assist the development of a multiple sequence alignment (MSA) of the *Plasmodium* protein kinase domains, each sequence was assessed using HMMER 3.3 <sup>20</sup> using the defined kinase families reported in Kinomer V1.0 <sup>23</sup>. Kinomer contains a profile for the AGC, CAMK, CK1, CMGC, RGC, STE, TKL and TYR ePK groups, and also includes a profile for the apicomplexan specific FIKK family <sup>12,24</sup>, but does not have a profile for the NEK family (traditionally considered to belong to the OPK group [see above]). In view of the importance of NEKs in all eukaryotes <sup>25</sup> including malaria parasites <sup>26</sup>, we designed a NEK profile with 21 known kinases from the NEK family <sup>27</sup> (see Methods for details). Using the amended Kinomer library (now containing a NEK profile), we performed a HMMER scan (hmmScan) of the kinomes and designated each sequence with the top hit based on score and E-value (see **Supplementary data 1** for full hmmScan

results). Each sequence was assessed to determine if it met the e-value thresholds for group assignment<sup>23</sup>. These threshold values differ for each group according to the highest E-value obtained during each group's hmm profile construction: AGC ( $2.7e^{-7}$ ), CAMK ( $3.2e^{-14}$ ), CK1 ( $3.2e^{-5}$ ), CMGC ( $1.2e^{-7}$ ), RGC ( $4.8e^{-5}$ ), STE ( $1.4e^{-6}$ ), TK ( $1.1e^{-9}$ ), TKL ( $1.7e^{-12}$ ) (see<sup>23</sup> for details). To allow for conservative group assignment we also considered assignments with a bit score <50 to be unreliable. This conservative value is double the default bit score threshold of the online HMMER tool [www.ebi.ac.uk/Tools/hmmer/](http://www.ebi.ac.uk/Tools/hmmer/). Phylogenetic trees with the *H. sapiens*, *P. falciparum* and *P. vivax* sequences for each of the ePK groups are available as **Supplementary Figures 2-8**.

**Multi-sequence alignment using the human kinome as a scaffold** – The human kinome comprises approximately 478 typical protein kinases, which contain a total of 497 typical protein kinase domains (as some sequences have two kinase domains)<sup>3</sup>. Of the 497 protein kinase domains known, over 270 have had their crystal structure solved. This is in stark comparison to the low number of solved protein kinase domains of *P. falciparum* and *P. vivax* (to date, 8 and 1 respectively). For *P. falciparum* these kinase domains are from PK5<sup>28</sup>, PK7<sup>29</sup>, PKG<sup>30</sup>, CDPK3<sup>31</sup>, CDPK4 (PDB: 4RGJ), MAPK2 (PDB: 3NIE), CK2<sup>32</sup> and CLK1 (PDB: 3LLT); for *P. vivax* only PKG<sup>30</sup>. The large number of kinase domain structures available supported the human kinome MSA, which was not possible for the *P. falciparum* and *P. vivax* kinomes. We therefore leveraged the homology between kinase families across species to aid in the MSA for both *P. falciparum* and *P. vivax* kinase domains (see Methods). The 17 conserved segments (230 amino acids) used for the alignment (see Methods) are depicted in **Figure 1** as sequence logos for the *Plasmodium* kinases, to highlight the conserved motifs important for kinase function from within the domain. Key motifs are (i) His-Arg-Asp (HRD), which is within the catalytic loop and stabilises the active site<sup>33</sup>, (ii) Asp-Phe-Gly (DFG), also within the activation loop and mediates allosteric conformational changes to regulate activation/inactivation of the enzyme<sup>34</sup>, and (iii) Ala-Pro-Glu (APE), which sits at the C-terminal end of the activation loop and stabilises the segment through docking to the domain's F-helix<sup>35</sup>. The HRD, DFG and APE motifs are part of the conserved segments in the catalytic loop (CL), Activation loop N-terminal (ALN) and Activation loop C-terminal (ALC), respectively (see **Figure 1**). To determine if the overall conserved segments of the kinase domain differ between *Plasmodium* and *Homo sapiens*, we generated sequence logos for the *Homo sapiens* sequences as well (**Supplementary Figure 9**). No large differences can be detected, but the consensus sequence is not as pervasive across the *Plasmodium* kinases, suggesting *Plasmodium* kinases can accommodate less stringent constraints at the primary structure level. However, this could be an artefact due to the kinomes of *Plasmodium* being smaller than that of humans, resulting in more divergent kinases making up a greater proportion of the kinome.

**Phylogenetic tree of the *Homo sapiens*, *P. falciparum* and *P. vivax* kinomes** – A phylogenetic tree (**Figure 2**) was constructed from the 230-column alignment consisting of the 17 aforementioned conserved segments (see Methods), as the removal of insertions improves accuracy<sup>36</sup>. To determine the phylogenetic relationship between the human kinome and that of *P. falciparum* and *P. vivax*, we included a total of 671 kinase domain sequences, covering all typical protein kinase sequences from these three organisms. The phylogeny relationships were determined using the RAxML GUI 2.0<sup>37,38</sup>. The definition of

the boundaries for each PK group was guided by the HMM profiles, the tree structure and the defined family assignment reported for each of the human kinases<sup>3</sup>. Six protein kinase previously flagged as orphans could now be confidently assigned to one of these ten groups, (see **Table 1** for changes in kinase group assignment). **Figure 3** depicts the number of kinases in each family per organism as a percentage of their kinome. This confirms previous reports that there are no *Plasmodium* kinases in either the TK or RGC groups<sup>12,15</sup>. However, we determined that both *P. falciparum* and *P. vivax* both have a single kinase belonging to the STE family (previously only reported in *P. vivax*<sup>39</sup>). In addition to the STEs, a clear reduction the AGC family (in comparison to the human kinome) can be observed as well. Orphan kinases make up a much larger percentage of the kinomes of both *Plasmodium* species (as compared to the human kinome), which presumably reflects the fact that the ePK groups were historically defined using Opisthokont organisms (metazoans and yeasts). The FIKK family, as previously reported, remains a prominent feature of the *P. falciparum* kinome, with a single member present in the *P. vivax*.

***Plasmodium* kinases with close homology to human kinases** – Each ePK group (including FIKK and orphan kinases) was assessed to determine if any particularly strong bootstrap support for homology between the *Plasmodium* and human protein kinases was observed. Surprisingly (in view of the divergent evolutionary paths of Alveolates and Metazoans), a number of noteworthy homologies were identified across most groups (**Supplementary data 2**). The *Plasmodium* kinases which exhibited bootstrap support to human kinase/s greater than 50 are listed in **Table 2**, along with their human homolog(s). **Table 2** notes that of the 16 *Plasmodium* kinases with bootstrap support above 50 to a human homolog(s), 10 of these belong to the CMGC group. Further, all *Plasmodium* kinases with bootstrap support values over 75, to a human homolog, belong to the CMGC group. Most notable are the *Plasmodium* kinase, cyclin-dependent-like kinase 3 (CLK3), serine/arginine protein kinase 1 (SRPK1), Casein kinase 2 alpha subunit (CK2a), Mitogen activated protein kinase 1 (MAPK1) and Glycogen synthase kinase 3 (GSK3) (**Figure 4**; a tree with the entire CMGC family, along with the CMGC kinases of *P. vivax*, is available in **Supplementary Figure 2**).

**CLK3** (PF3D7\_1114700), belongs to the CLK family of protein kinases, which in mammalian cells, facilitate phosphorylation of splicing factors<sup>40</sup>. In *Plasmodium*, *PfCLK3* is essential during asexual blood stage development<sup>41</sup>. *PfCLK3* has previously been assigned to the PRP4 subfamily of dual-specificity tyrosine-regulated kinases (DYRK)<sup>39</sup>. Our phylogenetic analysis confirms this finding, and further reports the striking homology between *PfCLK3* and *HsPRPF4B* (bootstrap support = 100).

**SRPK1** (PF3D7\_0302100), was initially considered to belong to the CLK family, however, it was reclassified as a SRPK following functional analysis<sup>42</sup>. SRPKs are closely related to the CLKs, and in mammalian cells, have a number of complex functions including mRNA processing and nuclear import (reviewed in<sup>43</sup>). In *Plasmodium*, *PfSRPK1* (previously known as *PfCLK4*) is essential during asexual blood stage development<sup>41</sup>. Here we can confirm the homology of *PfSRPK1* as the kinase clusters closely with the human SRPK1-3 (Bootstrap support = 99). Interestingly, *PfSRPK2* branches away before the SRPK and CLK division in the tree and is likely to have evolved from the precursor gene that gave rise to SRPKs and CLKs in the Opisthokont lineage. **CK2a** (PF3D7\_1108400), and **GSK3** (PF3D7\_0312400) are

present in all examined apicomplexan species. In *Plasmodium*, PfCK2 $\alpha$  and PfGSK3 are essential for blood stage development<sup>41</sup>. PfCK2 $\alpha$  has distinct homology to human CSNK2A1-3 (Bootstrap support = 99) and PfGSK3 shows strong homology to human GSK3 $\alpha/\beta$ . Interestingly, *Plasmodium* PK6, PK1, PF3D7\_1316000 and GSK3 cluster in the same branch as Human GSK $\alpha/\beta$ , which could suggest these three kinases derive from a common gene.

The Mitogen activation protein kinase family forms two relatively tight clusters of protein kinases within the CMGC group (**Figure 2**). MAPKs typically function as part of a three-tiered MAPK cascade, where a MAP3K phosphorylates a MAP2K which in turn phosphorylates a MAPK. PfMAPK1 (**Pfmap-1**, PF3D7\_1431500) clusters closely with human MAPK15 (ERK7, bootstrap support = 97), forming a clade that branches away from the majority of the MAPKs early in the tree (**Figure 2 and 4**). In humans, MAPK15 (ERK7) is an atypical MAPK that is activated by auto-phosphorylation rather than in the context of a classical 3-tier pathway<sup>44</sup>. In *P. falciparum* MAPK1 has been shown to be dispensable during erythrocytic development and for sporogony in the mosquito<sup>41,45</sup>. Curiously however, the other MAPK encoded by *P. falciparum*, MAPK2 (**pfmap-2**, PF3D7\_1113900) has been demonstrated to be elevated in MAPK1 knockouts, suggesting the parasite is able to adaptively compensate for reduced MAPK1 levels<sup>45</sup>. PfMAPK2 clusters within one of the primary branches of the human MAPK family (MAPK1/3/4/6/7 and NLK) (bootstrap support = 86). Further, within this family PfMAPK2 clusters closest to human MAPK1/3/4 and 6 (bootstrap support = 35, **Figure 4**). Despite a clear homolog to the above-mentioned family of MAPKs, *Plasmodium* does not possess a MAP2K orthologue to phosphorylate and activate PfMAPK2. In fact, both *P. falciparum* and *P. vivax* only encodes a single member of the STE group (containing the MAP2Ks), which does not cluster closely with any specific kinase (**Supplementary Figure 3**). Whether these enzymes function in pathways that implicate the MAPKs remains to be determined.

**Comparison of the *P. falciparum* and *P. vivax* kinome** – To directly compare the kinomes of *P. falciparum* and *P. vivax*, while preserving the phylogenetic tree structure, all *H. sapiens* branches were removed from the tree (**Figure 5**). As alluded to above, the kinomes of *P. vivax* and *P. falciparum*, despite their evolutionary distance, are very similar, with almost all kinases having a clear orthologue in the other species. There are only three distinct cases where no orthologue was observed (red arrows in **Figure 5**): first, as previously reported<sup>12</sup>, there is only one FIKK encoded by the *P. vivax* genome, versus a paralogous group of 21 sequences in *P. falciparum*. Second, PVP01\_0118800, which belongs to the TKL family, does not have an orthologue in *P. falciparum*. Third, PfCDPK2, from the CAMK family does not have an orthologue in *P. vivax*.

To understand which clades within the *Plasmodium* genus possess an orthologue of CDPK2, we aligned the kinase domains of all known CDPKs encoded by six distinct species; *P. falciparum*, *P. vivax*, *Plasmodium knowlesi*, *Plasmodium berghei*, *Plasmodium gaboni* and *Plasmodium gallinaceum* (see methods section). These sequences were assessed using RAXML<sup>37,38</sup> and a gene tree inferred from the results<sup>46</sup>. From the gene tree it is clear that CDPK2 is significantly different from its next closest homologue in the *Plasmodium* genus (CDPK3) (**Supplementary Figure 10**). To more extensively assess

which species in the genus *Plasmodium* encoded an orthologue of CDPK2 we completed a BLASTP search using the kinase domain of PfCDPK2. We compared the species that contained an orthologue of CDPK2 to the mitochondrial genome phylogeny for the *Plasmodium spp.* on PlasmoDB <sup>17</sup> (see method for details). This identified that the bird-infecting *Plasmodium gallinaceum* and *Plasmodium relictum*, as well as species in the *Laverania* lineage (*Plasmodium gaboni*, *Plasmodium rechenowi* and *P. falciparum*) all contain an orthologue of CDPK2, while species from the murine parasite clades and the other (non-Laveranian) primate-infecting parasites lineages do not (**Figure 6**). This is consistent with a whole-genome-based phylogeny suggesting that the *Laverania* have been founded by a single *Plasmodium* species switching from birds to African great apes (or vice versa, see below) , and suggest that CDPK2 has been lost in all other *Plasmodium* clades, or gained after the split between the clades <sup>47</sup>.

**Regulatory subunits of kinases** – Protein kinase regulatory subunits do not themselves have protein kinase activity, but are essential in the regulation of a select few protein kinases. Casein kinase 2 (CK2), which belongs to the CMGC group, forms a homo- or hetero-tetramer structure comprised of two regulatory subunits and two catalytic subunits <sup>48</sup>. *P. falciparum* encodes a single CK2 catalytic subunit (PF3D7\_1108400) and two different regulatory subunits (PF3D7\_1103700 and PF3D7\_1342400) <sup>32</sup>. BLASTP searches confirmed that *P. vivax* encodes orthologs to each of these subunits (CK2 catalytic subunit: PVP01\_0909200 and regulatory subunits: PVP01\_0904500, PVP01\_1212400) and no other. Protein Kinase A (PKA) belongs to the AGC ePK group, and, similar to CK2, the human holoenzyme is structured as a tetramer of two catalytic subunits and two regulatory subunits; cAMP binding to the regulatory subunits results in the release of the active catalytic subunits <sup>49</sup>. *P. falciparum* has previously been reported to encode a single PKA regulatory (PF3D7\_1223100) and a single catalytic subunit (PF3D7\_0934800) <sup>50</sup>, and the same is true for *P. vivax* (regulatory subunit: PVP01\_0733500; catalytic subunit: PVP01\_0733500). Lastly, cyclins are a diverse family of proteins which contain a conserved 5-helix bundled region known as the cyclin box, which enables binding to cyclin-dependent kinases (CDKs), stimulating their activity and hence playing a major role in cell cycle control <sup>51</sup>. *P. falciparum* encodes 3 readily identifiable cyclins, CYC1 (PF3D7\_1463700), CYC3 (PF3D7\_0518400) and CYC4 (PF3D7\_1304700) as well as CYC2 (PF3D7\_1227500), which appears to be more distantly related <sup>52,53</sup>. We completed an HMM search using both the PFAM IDs PF086134 (Cyclin) and PF00134 (Cyclin\_N) profiles, which revealed the presence of a fifth, previously unreported, putative cyclin in *P. falciparum* (PF3D7\_0605500). Additionally, the same search identified four possible cyclins in *P. vivax*, (PVP01\_1015500, PVP01\_1243100, PVP01\_1143400 and PVP01\_1405600). Phylogenetic analysis of the cyclin box (a highly conserved sequence among cyclins) showed that PVP01\_1243100, PVP01\_1015500 and PVP01\_1405600 are orthologous to PfCYC1, PfCYC3 and PfCYC4 respectively. Lastly, PVP01\_1143400 and PF3D7\_0605500, had not been identified previously and appear to be distantly related to Human Cyclin A (**Supplementary Figure 11**). Further refined phylogenetic analysis and functional validation of these putative cyclins are warranted.

## Concluding remarks

In this study we generated a complete human kinome comparison to *P. falciparum* and *P. vivax*, enabling the first comprehensive assessment of homologous protein kinases between an Apicomplexan parasite and its primary host. The striking kinome conservation observed across the evolutionarily distant species of *P. vivax* to *P. falciparum* together with previous studies of *P. berghei*<sup>54</sup> and earlier kinome assemblies of *P. falciparum* confirm that there is clear pressure for *Plasmodium spp.* to maintain the overwhelming majority of its kinome<sup>12,39</sup>. Though there are examples of clade-specific gene loss, such as CDPK2 and PVP01\_118800 reported here, the vast majority of kinases remain highly conserved. In *P. falciparum*, PfCDPK2 is critical for male gametogenesis<sup>55</sup>; therefore, it is likely that this function is fulfilled by another CDPK in the species where it is absent. The homology observed between CDPK2 and CDPK3 suggests that CDPK3 may fulfil this function, although this remains to be demonstrated (**Supplementary Figure 10**). PVP01\_118800 belongs to the TKL group and our kinome comparison indicates that it is most closely related to TKL4 of both *P. falciparum* and *P. vivax*, with a strong bootstrap support of 67 (**Figure 2 and 5**). A BLASTP search using PVP01\_118800's kinase domain as a query, indicated that an orthologue exists in all *Plasmodium* species with annotated genomes, except for species within the *Laverania* clade; this supports the hypothesis that the *Laverania* lineage results from a transfer from the bird-infecting parasites to the great apes, rather than the reverse<sup>47</sup>, and that TKL6 was lost after passage of the *Laverania* founding species from birds to great apes (**Figure 6**).

## Methods

**PIKK and Cyclin phylogeny** – The *Plasmodium falciparum* and *vivax*, PIKK and cyclin family of proteins were collected from PlasmoDB v50b<sup>17</sup> using the Pfam IDs of PF00134 and PF08613 (Cyclins) and PF00454 (PIKKs). To determine the validity of their assignment to the PIKK or Cyclin families, we created a HMM profile of the Cyclin box and atypical kinase domain (PIKKs) using a seed of the respective Pfam IDs available at pfam.xfam.org/<sup>56</sup>, as input into HMMER<sup>20</sup>. Following HMM profile identification the conserved sequences of the *Plasmodium* Cyclin box and atypical kinase domain of the PIKKs were aligned using ClustalOmega<sup>21</sup> and manually corrected in Jalview<sup>22</sup> before RAxML phylogenetic inference<sup>37,38</sup>. RAxML settings: maximum-likelihood rapid bootstrap with the PROTGAMMA substitution model LG4M, and AutoMRE. The gene tree was rendered using the webserver iTOL<sup>46</sup>.

**Plasmodium protein kinase sequences** – The *P. falciparum* sequences were collected from an earlier alignment<sup>1</sup>. *P. vivax* sequences were obtained by searching PlasmoDB v50b<sup>17</sup>. The predicted *P. vivax* proteome in PlasmoDB v50b was searched using the term “kinase”. The resultant list was further refined to only include, (i) sequences containing a Pfam ID of PF00069 or PF07714 (Protein kinase domain and protein tyrosine kinase, respectively) and (ii) sequences which contained the phrase “protein kinase” in the sequences annotated ‘Product description’ data.

**NEK HMM profile creation** – Kinomer<sup>23</sup>, the HMM profile databased used in this study did not include a profile for NEK family. The human protein kinase MSA which was the backbone of this study designated protein kinases into the NEK family. To be consistent with this and to enable comparison between

species, we developed a NEK profile using HMMER's hmmbuild function (HMMER 2.3.2)<sup>20</sup>, amending the Kinomer HMM profile with our NEK profile. Our NEK profile was comprised on 21 unique NEK sequences from 11 different organisms, which spanned the primary branches within the NEK family<sup>27</sup> (sequences available in **Supplementary Data 3**).

**Multiple Sequence Alignment** - The MSA developed for the human kinome contained 17 highly conserved blocks/segments interspersed with unaligned blocks of amino acids of variable length. The 17 aligned segments were defined using the structure of Human Aurora kinase A<sup>3</sup> and are similar to the regions initially defined by the original alignment published in 1988 by Hanks *et al.*<sup>57</sup>. These regions are named as follows; B1N, B1C, B2, B3, HC, B4, B5, HD, HE, CL, ALN, ALC, HF, FL, HG, HH and HI (see<sup>3</sup> for more details). Each of the *Plasmodium* kinase domains was aligned according to these 17 segments using the human MSA as a reference. For each *Plasmodium* kinase, the HMM profile was used to determine the closest related human kinases. For *Plasmodium* kinase domains designated as orphans, the nearest homologue was identified using PSI-BLAST and used to guide the alignment, leading to the definition of a 230- column amino acid sequence (one amino acid per column). The complete MSA contained a total of 3183 columns due to a number of large extensions in some *Plasmodium* kinases, notably SRPK1, PK1 and EST. The complete MSA of the *P. falciparum* and *P. vivax* kinases domains is available as **Supplementary data 4**.

**Phylogenetic relationship determination** - The phylogeny relationships were determined in the RAxML GUI 2.0<sup>37,38</sup>, using the 17 conserved sections of the MSA (230 columns) of the kinase domains of *H. sapiens*, *P. falciparum* and *P. vivax* as defined by<sup>3</sup>. The following parameters were used in RAxML: maximum-likelihood rapid bootstrap with the PROTgamma substitution model LG4M, and AutoMRE. A total of 200 bootstrapping runs was performed to meet the AutoMRE requirement. The gene tree was rendered using the webserver iTOL<sup>46</sup> before further annotations of the kinase families in Adobe illustrator. The definition of the borders for each family was guided by the HMM profiles, the tree structure and the defined family assignment reported for each of the human kinases<sup>3</sup>.

*Plasmodium* mitochondrial genome phylogeny – the mitochondrial genome of *P. gallinaceum*, *P. relictum*, *P. gaboni*, *P. rechenowi*, *P. falciparum*, *P. yoelii*, *P. berghei*, *P. chabaudi*, *P. vinckei*, *P. ovale*, *P. malariae*, *P. gonderi*, *P. fragile*, *P. coatneyi*, *P. knowlesi*, *P. vivax*, *P. Inui* and *P. cynomolgi* were obtained from GenBank<sup>58</sup> and aligned using ClustalOmega<sup>21</sup>, and imported into Jalview<sup>22</sup> and manually corrected. The MSA was imported into RAxML Gui 2.0<sup>38</sup>, and the following parameters input: maximum-likelihood rapid bootstrap with the gamma substitution model GTR with proportion of invariant sites through ML estimate (+I), with AutoMRE. A gene tree was inferred through the RAxML analysis using the 'best tree' and rendered with the interactive tree of life webserver (iTOL)<sup>46</sup>.

## List Of Abbreviations

ePK (eukaryotic protein kinase), aPK (atypical protein kinase), CAMK (Calcium/Calmodulin regulated kinases), CMGC (comprises the CDK, MAPK, GSK, and CLK families), CK1 (Casein Kinase I), NEK (NimA related kinases), AGC (comprises the protein kinase A/G/C families), RGC (Receptor Guanylate Cyclases), TKL (Tyrosine Kinase Like), TK (Tyrosine Kinase), ARK (Aurora Kinases), OPK (Orphan Protein Kinase) MSA (Multiple Sequence Alignment), CL, (Catalytic Loop), ALN (Activation Loop N-terminal) and ALC (Activation Loop C-terminal).

## Declarations

### Ethics approval and consent to participate

Not Applicable

### Consent for publication

N/A

### Authors contributions

J.A. and C.D. designed experiments. J.A. performed the experiments and analysed the results. J.A. and C.D. interpreted the data and wrote the manuscript. Both authors approved the manuscript.

### Availability of data and materials

All data are available in the Supplementary datasets provided.

### Competing interests

The authors declare no competing interests

### Funding and Acknowledgements

This work was supported by funding from the National Health and Medical Research Council, Australia , to J.A. and C.D. (APP2003712) and RMIT University (post-doctoral fellowship to J.A.).

## References

1. Adderley, J., Williamson, T. & Doerig, C. Parasite and Host Erythrocyte Kinomics of Plasmodium Infection. *Trends Parasitol* **37**, 508–524, doi:10.1016/j.pt.2021.01.002 (2021).
2. Wei, L. *et al.* Host-directed therapy, an untapped opportunity for antimalarial intervention. *Cell Rep Med* **2**, 100423, doi:10.1016/j.xcrm.2021.100423 (2021).
3. Modi, V. & Dunbrack, R. L., Jr. A Structurally-Validated Multiple Sequence Alignment of 497 Human Protein Kinase Domains. *Sci Rep* **9**, 19790, doi:10.1038/s41598-019-56499-4 (2019).

4. Dhiman, S. Are malaria elimination efforts on right track? An analysis of gains achieved and challenges ahead. *Infect Dis Poverty* **8**, 14, doi:10.1186/s40249-019-0524-x (2019).
5. Hamilton, W. L. *et al.* Evolution and expansion of multidrug-resistant malaria in southeast Asia: a genomic epidemiology study. *The Lancet. Infectious diseases* **19**, 943–951, doi:10.1016/S1473-3099(19)30392-5 (2019).
6. Menard, D. & Dondorp, A. Antimalarial Drug Resistance: A Threat to Malaria Elimination. *Cold Spring Harb Perspect Med* **7**, doi:10.1101/cshperspect.a025619 (2017).
7. Chaumont, C. *et al.* The SARS-CoV-2 crisis and its impact on neglected tropical diseases: Threat or opportunity? *PLoS Negl Trop Dis* **14**, e0008680, doi:10.1371/journal.pntd.0008680 (2020).
8. Zawawi, A. *et al.* The impact of COVID-19 pandemic on malaria elimination. *Parasite Epidemiol Control* **11**, e00187, doi:10.1016/j.parepi.2020.e00187 (2020).
9. Arora, N., L, C. A. & Pannu, A. K. Towards Eradication of Malaria: Is the WHO's RTS,S/AS01 Vaccination Effective Enough? *Risk Manag Healthc Policy* **14**, 1033–1039, doi:10.2147/RMHP.S219294 (2021).
10. Roskoski, R., Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2021 update. *Pharmacol Res* **165**, 105463, doi:10.1016/j.phrs.2021.105463 (2021).
11. Taylor, S. S., Keshwani, M. M., Steichen, J. M. & Kornev, A. P. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos Trans R Soc Lond B Biol Sci* **367**, 2517–2528, doi:10.1098/rstb.2012.0054 (2012).
12. Ward, P., Equinet, L., Packer, J. & Doerig, C. Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. *BMC Genomics* **5**, 79, doi:10.1186/1471-2164-5-79 (2004).
13. Hanks, S. K. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol* **4**, 111, doi:10.1186/gb-2003-4-5-111 (2003).
14. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934, doi:10.1126/science.1075762 (2002).
15. Talevich, E., Tobin, A. B., Kannan, N. & Doerig, C. An evolutionary perspective on the kinome of malaria parasites. *Philos Trans R Soc Lond B Biol Sci* **367**, 2607–2618, doi:10.1098/rstb.2012.0014 (2012).
16. Miranda-Saavedra, D., Gabaldon, T., Barton, G. J., Langsley, G. & Doerig, C. The kinomes of apicomplexan parasites. *Microbes Infect* **14**, 796–810, doi:10.1016/j.micinf.2012.04.007 (2012).
17. Aurrecochea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* **37**, D539-543, doi:10.1093/nar/gkn814 (2009).
18. de Castro, E. *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34**, W362-365, doi:10.1093/nar/gkl124 (2006).

19. Holland, Z., Prudent, R., Reiser, J. B., Cochet, C. & Doerig, C. Functional analysis of protein kinase CK2 of the human malaria parasite *Plasmodium falciparum*. *Eukaryot Cell* **8**, 388–397, doi:10.1128/EC.00334-08 (2009).
20. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29-37, doi:10.1093/nar/gkr367 (2011).
21. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539, doi:10.1038/msb.2011.75 (2011).
22. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191, doi:10.1093/bioinformatics/btp033 (2009).
23. Martin, D. M., Miranda-Saavedra, D. & Barton, G. J. Kinomer v. 1.0: a database of systematically classified eukaryotic protein kinases. *Nucleic Acids Res* **37**, D244-250, doi:10.1093/nar/gkn834 (2009).
24. Schneider, A. G. & Mercereau-Puijalon, O. A new Apicomplexa-specific protein kinase family: multiple members in *Plasmodium falciparum*, all with an export signature. *BMC Genomics* **6**, 30, doi:10.1186/1471-2164-6-30 (2005).
25. Fry, A. M., Bayliss, R. & Roig, J. Mitotic Regulation by NEK Kinase Networks. *Front Cell Dev Biol* **5**, 102, doi:10.3389/fcell.2017.00102 (2017).
26. Carvalho, T. G., Doerig, C. & Reininger, L. Nima- and Aurora-related kinases of malaria parasites. *Biochim Biophys Acta* **1834**, 1336–1345, doi:10.1016/j.bbapap.2013.02.022 (2013).
27. Parker, J. D., Bradley, B. A., Mooers, A. O. & Quarmby, L. M. Phylogenetic analysis of the Neks reveals early diversification of ciliary-cell cycle kinases. *PLoS One* **2**, e1076, doi:10.1371/journal.pone.0001076 (2007).
28. Holton, S. *et al.* Structures of *P. falciparum* PfPK5 test the CDK regulation paradigm and suggest mechanisms of small molecule inhibition. *Structure* **11**, 1329–1337, doi:10.1016/j.str.2003.09.020 (2003).
29. Merckx, A. *et al.* Structures of *P. falciparum* protein kinase 7 identify an activation motif and leads for inhibitor design. *Structure* **16**, 228–238, doi:10.1016/j.str.2007.11.014 (2008).
30. El Bakkouri, M. *et al.* Structures of the cGMP-dependent protein kinase in malaria parasites reveal a unique structural relay mechanism for activation. *Proc Natl Acad Sci U S A* **116**, 14164–14173, doi:10.1073/pnas.1905558116 (2019).
31. Wernimont, A. K. *et al.* Structures of parasitic CDPK domains point to a common mechanism of activation. *Proteins* **79**, 803–820, doi:10.1002/prot.22919 (2011).
32. Ruiz-Carrillo, D. *et al.* The protein kinase CK2 catalytic domain from *Plasmodium falciparum*: crystal structure, tyrosine kinase activity and inhibition. *Sci Rep* **8**, 7365, doi:10.1038/s41598-018-25738-5 (2018).
33. Strong, T. C., Kaur, G. & Thomas, J. H. Mutations in the catalytic loop HRD motif alter the activity and function of *Drosophila* Src64. *PLoS One* **6**, e28100, doi:10.1371/journal.pone.0028100 (2011).

34. Treiber, D. K. & Shah, N. P. Ins and outs of kinase DFG motifs. *Chem Biol* **20**, 745–746, doi:10.1016/j.chembiol.2013.06.001 (2013).
35. Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem Sci* **36**, 65–77, doi:10.1016/j.tibs.2010.09.006 (2011).
36. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564–577, doi:10.1080/10635150701472164 (2007).
37. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, doi:10.1093/bioinformatics/btu033 (2014).
38. Edler, D., Klein, J., Antonelli, A. & Silvestro, D. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods in Ecology and Evolution* **12**, 373–377, doi:https://doi.org/10.1111/2041-210X.13512 (2021).
39. Talevich, E., Mirza, A. & Kannan, N. Structural and evolutionary divergence of eukaryotic protein kinases in Apicomplexa. *BMC Evol Biol* **11**, 321, doi:10.1186/1471-2148-11-321 (2011).
40. Zhou, Z. & Fu, X. D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **122**, 191–207, doi:10.1007/s00412-013-0407-z (2013).
41. Solyakov, L. *et al.* Global kinomic and phospho-proteomic analyses of the human malaria parasite *Plasmodium falciparum*. *Nat Commun* **2**, 565, doi:10.1038/ncomms1558 (2011).
42. Dixit, A., Singh, P. K., Sharma, G. P., Malhotra, P. & Sharma, P. PfSRPK1, a novel splicing-related kinase from *Plasmodium falciparum*. *J Biol Chem* **285**, 38315–38323, doi:10.1074/jbc.M110.119255 (2010).
43. Giannakouros, T., Nikolakaki, E., Mylonis, I. & Georgatsou, E. Serine-arginine protein kinases: a small protein kinase family with a large cellular presence. *FEBS J* **278**, 570–586, doi:10.1111/j.1742-4658.2010.07987.x (2011).
44. Abe, M. K. *et al.* ERK7 is an autoactivated member of the MAPK family. *J Biol Chem* **276**, 21272–21279, doi:10.1074/jbc.M100026200 (2001).
45. Dorin-Semlat, D. *et al.* Functional characterization of both MAP kinases of the human malaria parasite *Plasmodium falciparum* by reverse genetics. *Mol Microbiol* **65**, 1170–1180, doi:10.1111/j.1365-2958.2007.05859.x (2007).
46. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293–W296, doi:10.1093/nar/gkab301 (2021).
47. Pick, C., Ebersberger, I., Spielmann, T., Bruchhaus, I. & Burmester, T. Phylogenomic analyses of malaria parasites and evolution of their exported proteins. *BMC Evol Biol* **11**, 167, doi:10.1186/1471-2148-11-167 (2011).
48. Niefind, K., Guerra, B., Ermakowa, I. & Issinger, O. G. Crystal structure of human protein kinase CK2: insights into basic properties of the CK2 holoenzyme. *EMBO J* **20**, 5320–5331, doi:10.1093/emboj/20.19.5320 (2001).

49. Turnham, R. E. & Scott, J. D. Protein kinase A catalytic subunit isoform PRKACA; History, function and physiology. *Gene* **577**, 101–108, doi:10.1016/j.gene.2015.11.052 (2016).
50. Merckx, A. *et al.* Plasmodium falciparum regulatory subunit of cAMP-dependent PKA and anion channel conductance. *PLoS Pathog* **4**, e19, doi:10.1371/journal.ppat.0040019 (2008).
51. Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* **13**, 261–291, doi:10.1146/annurev.cellbio.13.1.261 (1997).
52. Roques, M. *et al.* Plasmodium P-Type Cyclin CYC3 Modulates Endomitotic Growth during Oocyst Development in Mosquitoes. *PLoS Pathog* **11**, e1005273, doi:10.1371/journal.ppat.1005273 (2015).
53. Merckx, A. *et al.* Identification and initial characterization of three novel cyclin-related proteins of the human malaria parasite Plasmodium falciparum. *J Biol Chem* **278**, 39839–39850, doi:10.1074/jbc.M301625200 (2003).
54. Tewari, R. *et al.* The systematic functional analysis of Plasmodium protein kinases identifies essential regulators of mosquito transmission. *Cell Host Microbe* **8**, 377–387, doi:10.1016/j.chom.2010.09.006 (2010).
55. Hitz, E., Balestra, A. C., Brochet, M. & Voss, T. S. PfMAP-2 is essential for male gametogenesis in the malaria parasite Plasmodium falciparum. *Sci Rep* **10**, 11930, doi:10.1038/s41598-020-68717-5 (2020).
56. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28**, 405–420, doi:10.1002/(sici)1097-0134(199707)28:3<405::aid-prot10>3.0.co;2-l (1997).
57. Hanks, S. K., Quinn, A. M. & Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52, doi:10.1126/science.3291115 (1988).
58. Sayers, E. W. *et al.* GenBank. *Nucleic Acids Res* **49**, D92-D96, doi:10.1093/nar/gkaa1023 (2021).
59. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190, doi:10.1101/gr.849004 (2004).

## Tables

**Table 1 – Group assignment of previously orphan *P. falciparum* kinases**

Kinase ID	Name	Group Assignment (this study)
PF3D7_0309200	ARK2	ARK
PF3D7_1356800	ARK3	ARK
PF3D7_1121900	-	AGC
PF3D7_0605300	ARK1	ARK
PF3D7_1145200	-	CAMK
PF3D7_1441300	-	CAMK
PF3D7_0213400	PK7	CAMK
PF3D7_0926000	-	TKL
PF3D7_0926100	-	TKL

Table 2 – List of *Plasmodium* kinases with bootstrap support greater than 50 to a human homolog

<i>P. falciparum</i> ID	<i>P. vivax</i> ID	Name	Kinase group	closest human homolog/s	bootstrap support
PF3D7_1114700	PVP01_0915400	CLK3	CMGC	PRPF4B	100
PF3D7_0302100	PVP01_0302100	SRPK1	CMGC	SRPK1, SRPK2, SRPK3	99
PF3D7_1108400	PVP01_0909200	CK2a	CMGC	CSNK2A1, CSNK2A2, CSNK2A3	99
PF3D7_1431500	PVP01_1431500	MAPK1	CMGC	MAPK15	97
PF3D7_0312400	PVP01_0827800	GSK3	CMGC	GSK3A, GSK3B	87
PF3D7_1356900	PVP01_1115000	PK5	CMGC	CDK5, CDK14, CDK15, CDK16, CDK17, CDK18	84
PF3D7_0708300	PVP01_0108600	BUD32	Orphan	TP53RK	68
PF3D7_1121300	PVP01_0921900	TKL2	TKL	IRAK1, IRAK2, IRAK3, IRAK4	68
PF3D7_1228300	PVP01_1446500	NEK1	NEK	NEK2	67
PF3D7_1445400	PVP01_1261600	CLK1	CMGC	CLK1, CLK2, CLK3, CLK4	61
PF3D7_0415300	PVP01_0521800	CRK3	CMGC	CDK9, CDK12, CDK13	60
PF3D7_0821100	PVP01_0513200	PK1	CMGC	GSK3A, GSK3B	57
PF3D7_1316000	PVP01_1416900	-	CMGC	GSK3A, GSK3B	57
PF3D7_1246900	PVP01_1464000	PKB	AGC	SGK1, SGK2, SGK3, AKT1, AKT2, AKT3, PRKCI, PRKCZ, PRKCH, PRKCE, PRKCB, PRKCG, PRKCA, PRKCQ, PRKCD, PKN1, PKN2, PKN3	56
PF3D7_1136500	PVP01_0937400	CK1	CK1	CSNK1A1, CSNK1A1L	55
PF3D7_1436600	PVP01_1312900	PKG	AGC	PRKG1, PRKG2	53

## Figures

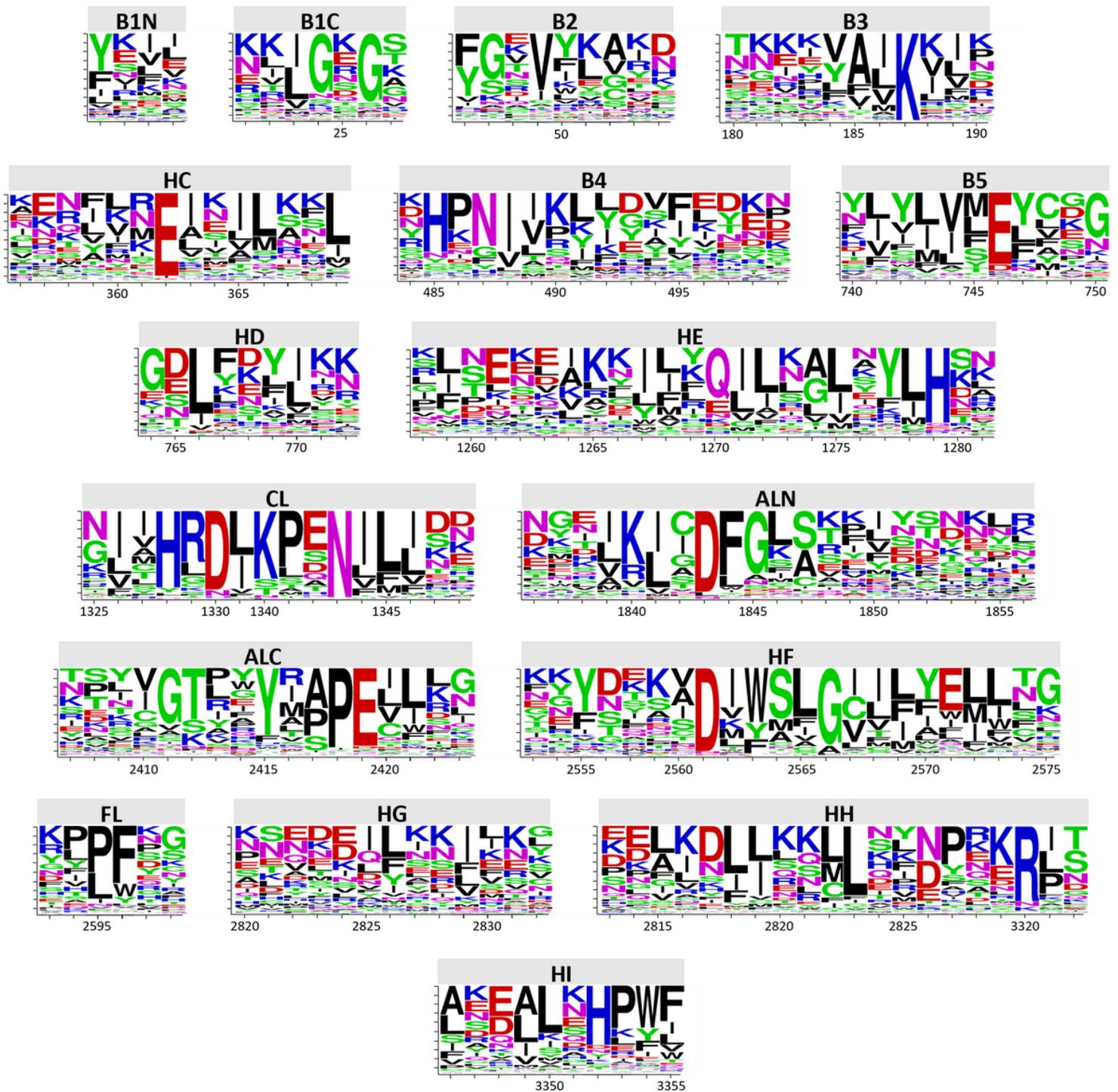


Figure 1

Sequence logos of the conserved regions in the multisequence alignment for the kinase domains of *P. falciparum* and *P. vivax*. Aligned regions of the kinase domain defined by <sup>39</sup> and logo generated using the webserver WebLogo (<https://weblogo.berkeley.edu/>) <sup>59</sup> and edited in Adobe Illustrator.

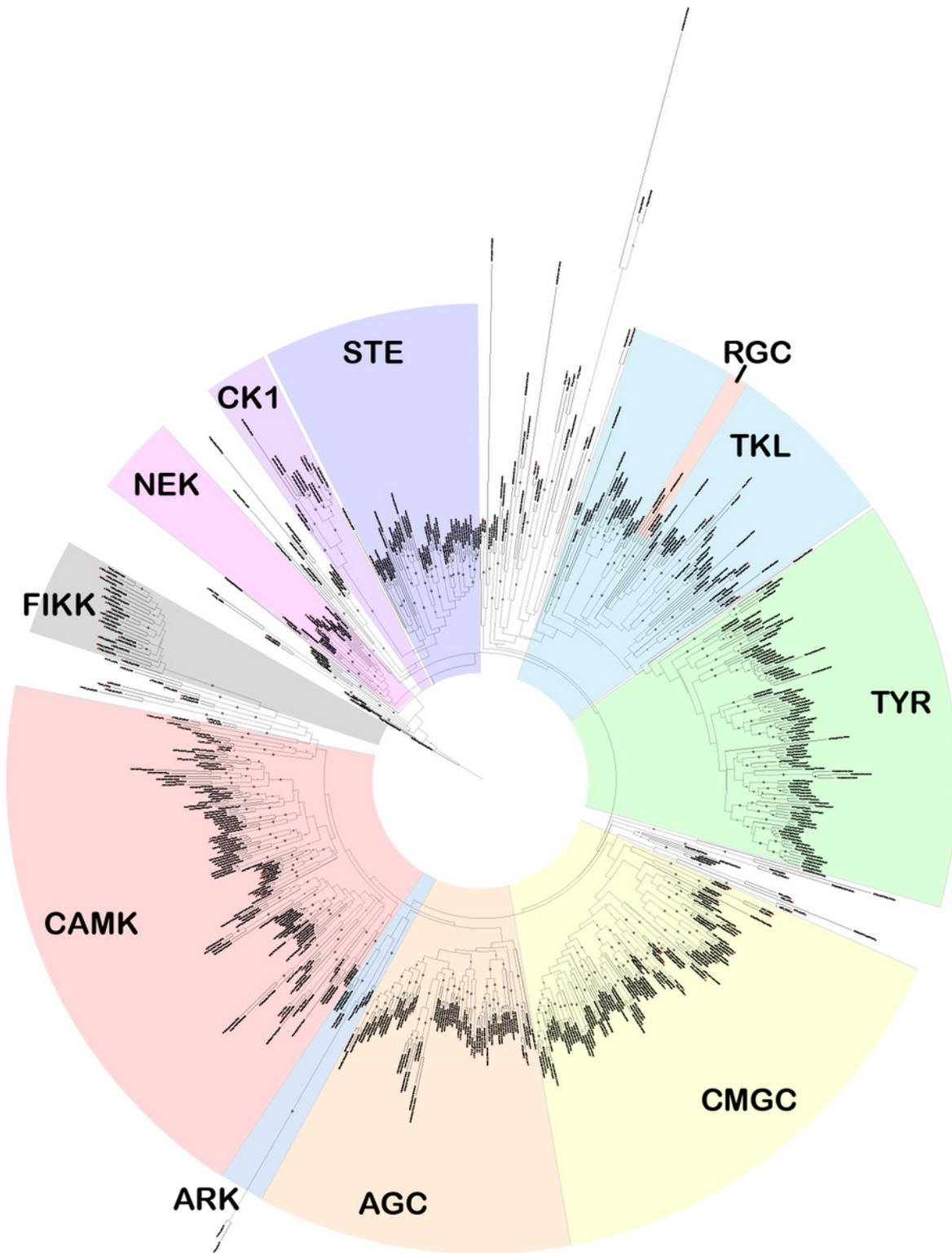


Figure 2

Phylogenetic tree containing the protein kinome of *Homo sapiens*, *Plasmodium falciparum*, *Plasmodium vivax*. The tree is represented in a circular format and contains a total of 671 protein kinases sequences excluding the atypical kinases (*H. sapiens* - 497, *P. falciparum* - 98 and *P. vivax* - 78.). *P. falciparum* sequences were accessed from <sup>1</sup>, *H. sapiens* sequences were accessed from <sup>3</sup> and *P. vivax* sequences were identified from PlasmoDB <sup>17</sup> and initially aligned using ClustalOmega <sup>21</sup>. All of the kinase

sequences were imported into Jalview<sup>22</sup>. Using the human kinases as a template, the *Plasmodium* kinases were aligned into the conserved regions as defined by<sup>3</sup>. The resultant 230 column alignment was assessed using RAxML to infer phylogenetic distances and determine bootstrapping<sup>37</sup>. RAxML Gui 2.0<sup>38</sup> was used with the following parameters: maximum-likelihood rapid bootstrap with the PROTgamma substitution model LG4M, with AutoMRE. A gene tree was inferred through the RAxML analysis using the 'best tree' and rendered with the interactive tree of life webserver (iTOL)<sup>46</sup>. HMMER profiling was performed for *P. falciparum* and *P. vivax* kinases using the defined families of Kinomer<sup>23</sup> with the addition of the NEK family (see Supplementary data 3). Using the tree structure, HMMER results and the known human kinase family assignments the *Plasmodium* kinases were assigned to the 9 typical protein kinase groups. These family assignments were annotated using Adobe illustrator along with the Aurora kinase family (ARK) and the Apicomplexan-specific kinase family FIKK. Orphan, or 'other' kinases are largely unassigned to families (white background). Bootstraps values above 50 are represented as circles on the associated branches, larger circles indicate higher bootstrap values. *Plasmodium* kinases are indicated with a Red star.

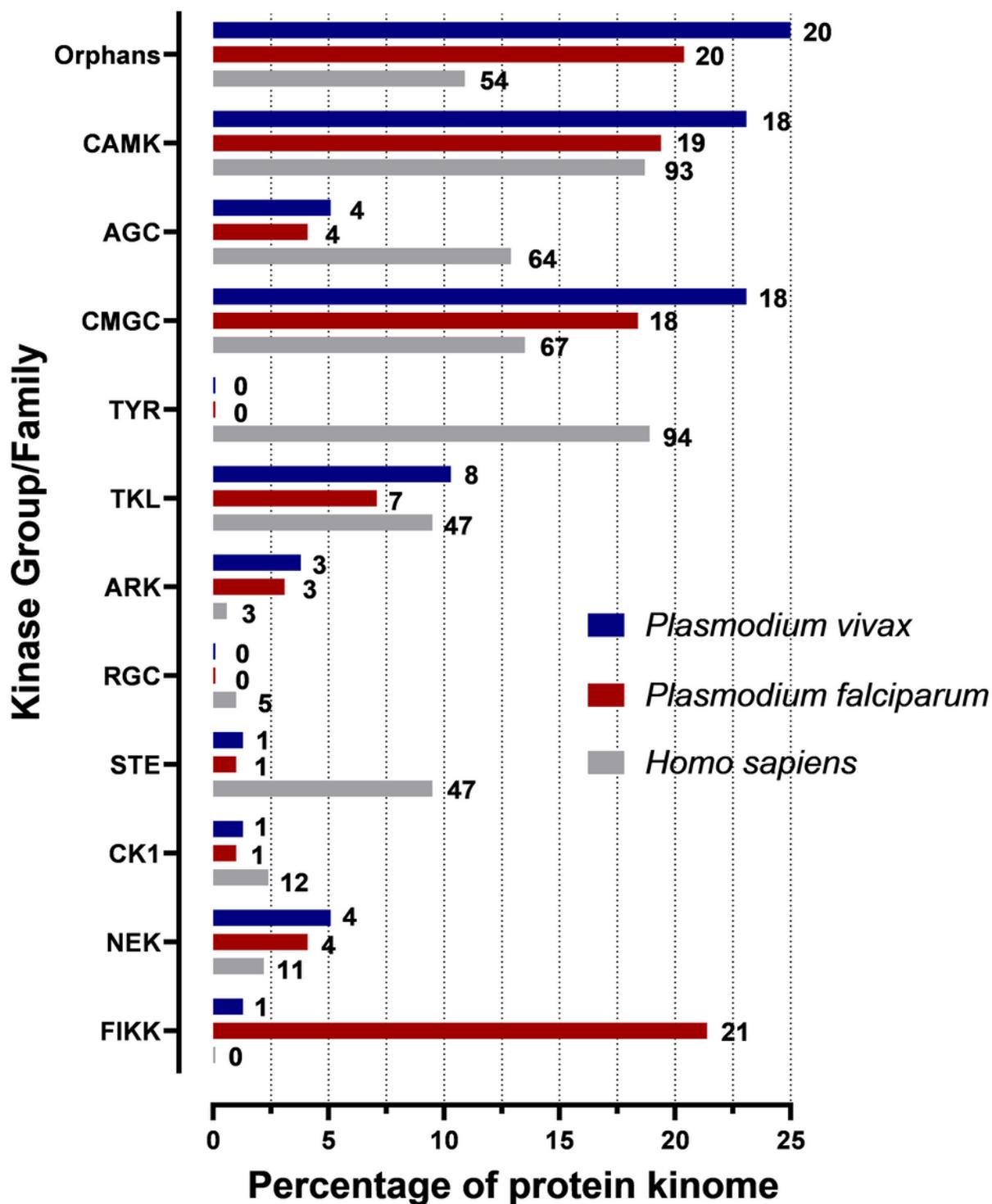
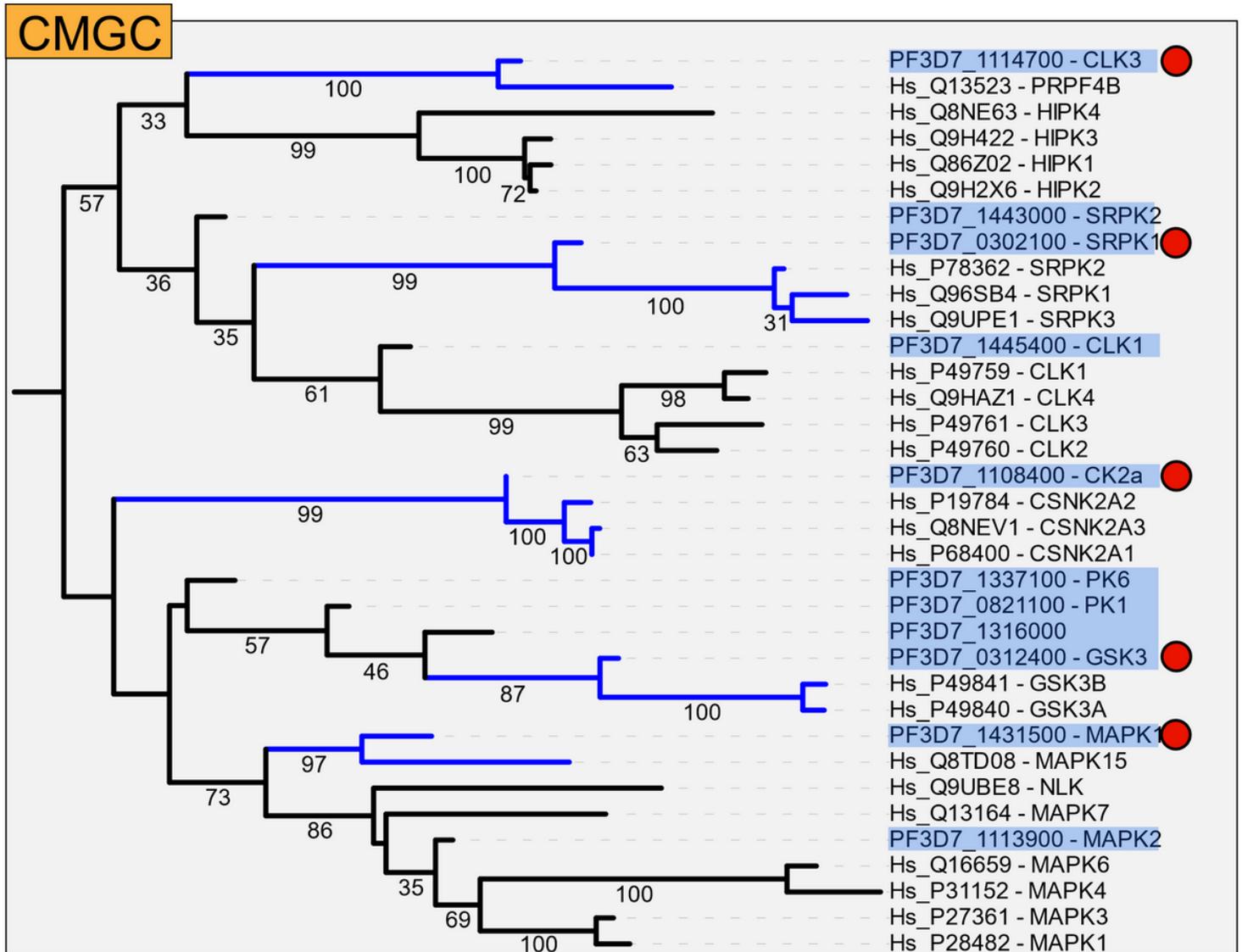


Figure 3

Visualisation of the protein kinase family membership across *Homo sapiens*, *Plasmodium falciparum*, *Plasmodium vivax*. The nine typical protein kinase families along with the Aurora kinase family (ARK) and the Apicomplexan-specific family of FIKK were included here. The remaining unassigned kinases denoted as Orphans. Each group/family is represented as a percentage of the total protein kinome for each organism. Note atypical protein kinases are not included in this analysis. The number next to each

bar indicates the number of kinases which belong to each of the respective families for each organism. Blue = *Plasmodium vivax*, Red = *Plasmodium falciparum* and Grey = *Homo sapiens*.



**Figure 4**

**Truncated phylogenetic tree of the CMGC group, focused on branches with high homology between *Plasmodium* and human sequences.** A small number of *Plasmodium* kinases within the CMGC group exhibited high homology to human kinases. These kinases were *Pf*CLK3, *Pf*SRPK1, *Pf*MAPK1, *Pf*CKa and *Pf*GSK3 (denoted with a red circle). The associated strong bootstrap support (>80) has been coloured blue, along with the branches to the human homologs. Bootstrap values are listed on the associated branches (when >30) *Plasmodium* kinases are highlighted in blue.

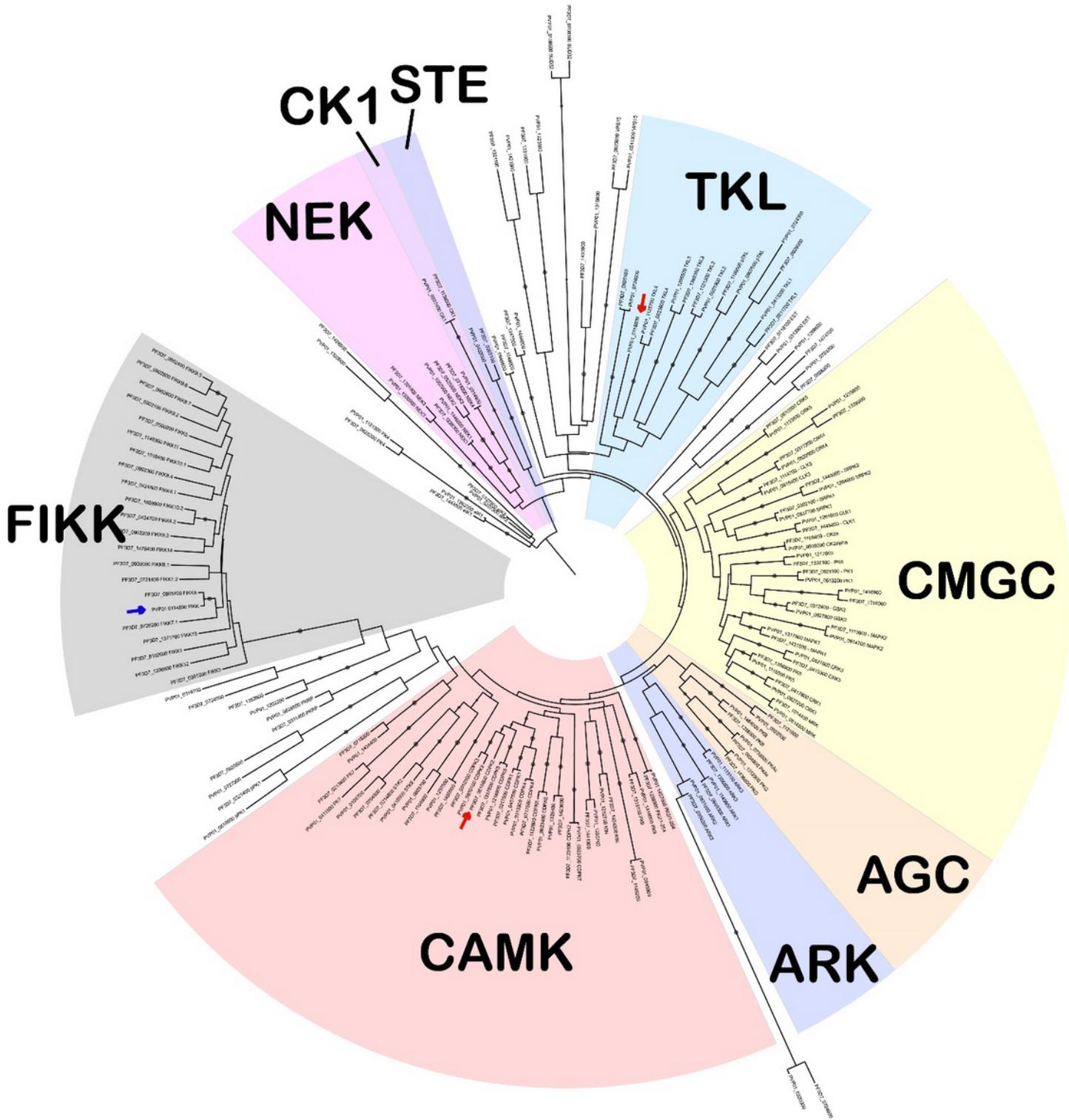
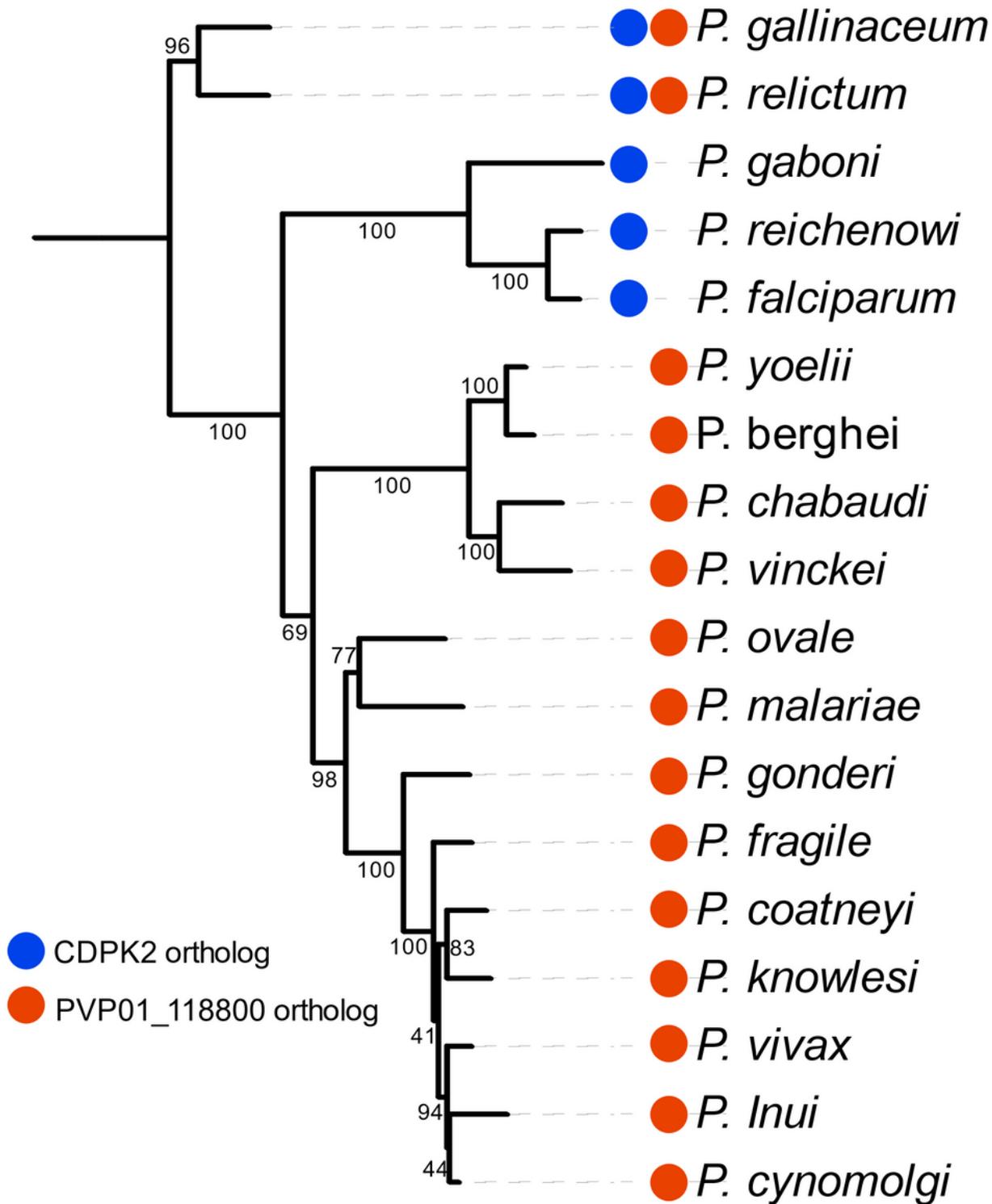


Figure 5

**Comparative Phylogenetic tree of *Plasmodium vivax* and *Plasmodium falciparum*.** Phylogenetic tree indicating the kinases shared and unique to each of the two species. Red arrow indicates the kinases PVP01\_011880 and PF3D7\_061060 (CDPK2) which do not have an equivalent kinase in the other species. Blue arrow indicates PVP01\_011480 FIKK the only member of the FIKK family that *P. vivax* encodes. See Figure 1 legend for details regarding the assembly and construction of the phylogenetic

tree. Typical protein group families were annotated using Adobe illustrator along with the Aurora kinase family (ARK) and the Apicomplexan-specific kinase family FIKK. Orphan, or 'other' kinases are largely unassigned to families (white background). Bootstraps values above 50 are represented as circles on the associated branches, larger circles indicate higher bootstrap values, note these values relate to the full tree illustrated in Figure 2.



## Figure 6

**Phylogenetic tree of *Plasmodium spp.* illustrating CDPK2 and PVP01\_118800 orthologs.** Tree assembled using the mitochondrial genomes of each species (see methods for details). A gene tree was inferred through the RAxML analysis using the 'best tree' and rendered with the tree of life webserver (iTOL) <sup>46</sup>. CDPK2 and PVP01\_118800 orthologs were identified through blast searches using the kinase domains. *Plasmodium* species which have orthologs to CDPK2 are indicated by the blue circle, while species with an ortholog of PVP01\_118800 are indicated by the red circle. Bootstrap support for the gene tree is indicated on each of the branches, only values above 40 are displayed.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfigure1PIKKtree.pdf](#)
- [Supplementaryfigure2CMGCtree.pdf](#)
- [Supplementaryfigure3STETree.pdf](#)
- [Supplementaryfigure4TKLtree.pdf](#)
- [Supplementaryfigure5CK1andNEKtrees.pdf](#)
- [Supplementaryfigure7ARKtree.pdf](#)
- [Supplementaryfigure8AGCtree.pdf](#)
- [Supplementaryfigure9Kinasedomainsequencelogos.pdf](#)
- [Supplementaryfigure10CDPKtree.pdf](#)
- [Supplementaryfigure11Cyclintree.pdf](#)
- [Supplementaryfileinformationandlegends.docx](#)
- [Supplementarydata1PfPvHMMResults.txt](#)
- [Supplementarydata2Plasmodiumhomologytohumankinases.xlsx](#)
- [Supplementarydata3NEKSequences.sto](#)
- [Supplementarydata4MSAofHsPfPvkinasedomains.fa](#)