

PhyloPrimer manual

Contents

| | |
|--|-----------|
| 1 This manual | 1 |
| 2 The web platform | 2 |
| 2.1 Home page | 2 |
| 2.2 Input pages | 2 |
| 2.2.1 Dynamic selection | 3 |
| 2.2.2 Premade selection | 9 |
| 2.2.3 Oligo check | 10 |
| 2.3 Oligo pages | 10 |
| 2.3.1 Primer Design | 10 |
| 2.3.2 Primer and Probe Design | 14 |
| 2.3.3 Single Oligo Design | 15 |
| 2.4 Result page | 16 |
| 2.4.1 Specifics | 17 |
| 2.4.2 ΔG | 17 |
| 2.4.3 GenBank BLAST | 17 |
| 2.4.4 Your BLAST | 18 |
| 2.4.5 Archive | 18 |
| 3 Input formats | 21 |
| 3.1 DNA sequences | 21 |
| 3.2 DNA alignments | 22 |
| 3.3 Newick tree | 22 |
| 3.4 Tab-delimited tables | 22 |
| 3.5 Oligo assays | 23 |
| 4 Oligo Design and Scoring System | 23 |
| 5 Consensus calculation | 27 |
| 6 Thermodynamic calculation | 27 |
| 6.1 Melting temperature | 27 |
| 6.2 ΔG calculation | 30 |
| 6.3 Formula corrections | 31 |
| 6.3.1 Monovalent ions and Mg^{2+} | 33 |
| 6.3.2 Salt-corrected melting temperature | 34 |
| 6.3.3 Salt-corrected ΔG | 36 |
| 7 Database | 36 |
| 8 Server Specifics and Coding language | 37 |
| Appendices | 38 |

1 This manual

PhyloPrimer was born from a collaboration between the University of Bristol and the bioremediation company ENOVEO. The idea was to create a user-friendly platform to i) retrieve DNA sequences similar to a gene of interest, ii) tree-based select and design oligos (e.g. primer and probes) suitable for different PCR applications, iii)

check the oligos for the presence of secondary structures and iv) study each oligo with in silico tests to determine which organisms they can potentially target. The strength of PhyloPrimer derives from the possibility of retrieving the DNA sequences that will be used for the oligo design through a dynamic tree. This functionality makes this software a powerful tool for designing taxon specific oligos. PhyloPrimer designs the oligos uniquely on the sequences selected from the dynamic tree and tests all the oligos for their specificity to the target organisms. The software can also be used for designing oligos which target a gene without any taxonomic specificity, for designing oligos from preselected sequences and for checking predesigned oligos. You will find more details on what this software can do in the manual. Now just a few words to clarify what PhyloPrimer is not:

- it is not a tool for the design of eukaryotic oligos. PhyloPrimer was developed with the study of prokaryotes in mind, it does not deal with intron and exon regions so it is not advised to use this software with anything other than prokaryotic data. Furthermore, the dynamic tree is built only with prokaryotic sequences.
- it is not a tool for the design of degenerate oligos specifically. PhyloPrimer uses a consensus approach and it designs the oligos from a consensus sequence calculated from a DNA alignment. Therefore it will not introduce degeneracy on purpose and will design oligos containing degenerate bases only if present in the consensus sequence and if necessary to the design of suitable oligos.
- it is not a tool for phylogenetic tree exploration. The dynamic tree is only constructed to help the user to select the DNA sequences that will be used for the consensus construction. Therefore the tree's purpose is only to show the sequence grouping and not any evolutionary relationship between the different clusters as no accurate phylogenetic tree construction algorithms or bootstraps are used. Furthermore it is not advisable to upload much data just for the sake of data exploration as the software slows down considerably when a lot of data is uploaded.
- it is not a magic tool. Please do know the gene family you are studying and set the oligo design parameters considering your gene sequence properties.

PhyloPrimer is available through a web platform at <https://www.cerealsdb.uk.net/cerealgenomics/phyloprimer/> and it is free to use. With this manual we want to provide a guide for the use of PhyloPrimer and to provide a description of the processes happening in the background. Examples of applications of PhyloPrimer can be found in our paper: XXXX.

2 The web platform

The PhyloPrimer web platform is structured with sequential web pages that can be categorized into four different groups: i) the home page, ii) the input pages, iii) the oligo pages and iv) the result page. From the home page, the user can select one of the three different input pages available for uploading the data (e.g. DNA sequences, DNA alignments and newick tree) where each page corresponds to a different modality to use PhyloPrimer. Once the data are uploaded, the user is redirected to the oligo pages where there are different parameter settings for designing either primer assays, primer and probe assays or single oligos. Once the user submits these parameters, the oligo design and the oligo check are performed on the web server. As soon as PhyloPrimer has finished the analyses, the user will receive an email with a link to the Result Page where the user will be able to explore the designed oligos and select the ones will be used for future work (Figure 1).

2.1 Home page

The home page reports a brief description of what PhyloPrimer is and, clicking either on the button **Dynamic selection**, **Premade selection** or **Oligo check**, the user is redirected to the appropriate page for the uploading of the data PhyloPrimer will use for the oligo design. In this page, the user can also find information on the last PhyloPrimer's database upgrade (Section 7).

2.2 Input pages

The three different input pages are only selectable from the PhyloPrimer home page. The choice of which input page to use is straightforward. If the user wants to amplify a gene or region of interest and wants to look for similar DNA sequences before proceeding to the primer design, the **Dynamic selection** input page should be used. If the user has already selected the DNA sequences to use for the primer design, the right choice is the **Premade selection** input page. If the user wants to check predesigned oligos, the **Oligo check** page should be used. Depending on which input page is selected and which input is uploaded, PhyloPrimer will start different

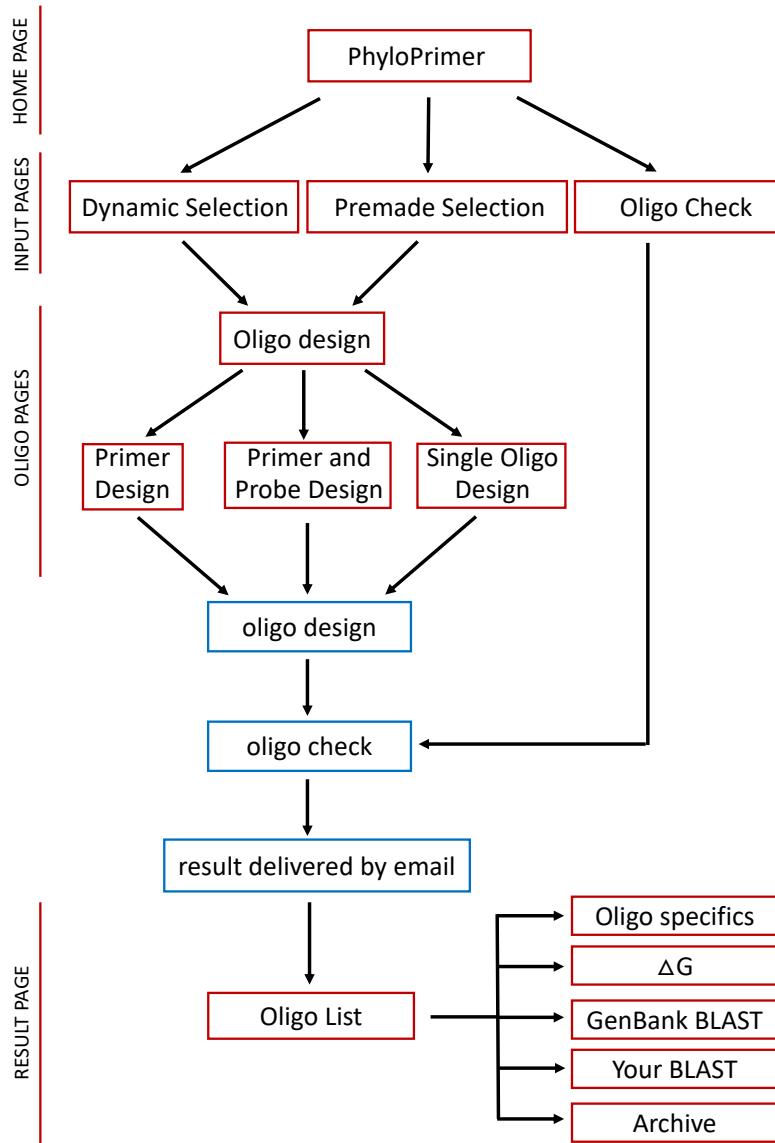


Figure 1. PhyloPrimer structure where the red boxes indicate the web pages whereas the blue boxes indicate the processes that are run on the server side.

processes on the server side (Figure 2). If in doubt, the input formats can be checked by clicking on the blue links in any of the web page. Also, any time that the user changes an input field value, this will turn blue.

2.2.1 Dynamic selection

The first mode, **Dynamic selection**, is used when the user wants to select DNA sequences, similar to a starting gene or a DNA fragment of interest, that will be aligned by PhyloPrimer to calculate the consensus sequence to be used for the oligo design. This page is divided in four sections named Inputs, Uploads, Dynamic Tree and Consensus. When the **Dynamic selection** page is first loaded, the Inputs section is the only visible one (Figure 3). This section has three radio buttons, each representing a different input option. Selecting a specific radio button enables its input fields and disables the others. In Figure 3, for example, the first field is active and a fasta file containing one gene of interest has been uploaded. The user can upload either up to ten genes/DNA regions of interest (Figure 3a), up to 500 DNA sequences (Figure 3b) or a newick tree (with a maximum of 500 tree leaves, Figure 3c):

- Up to 10 DNA sequences. In case this option is selected, PhyloPrimer will first perform a BLAST search against a subset of the GenBank database containing only microbial sequences (more details in Section 7), then will run a MAFFT alignment and finally output a dynamic tree representing all the input sequences

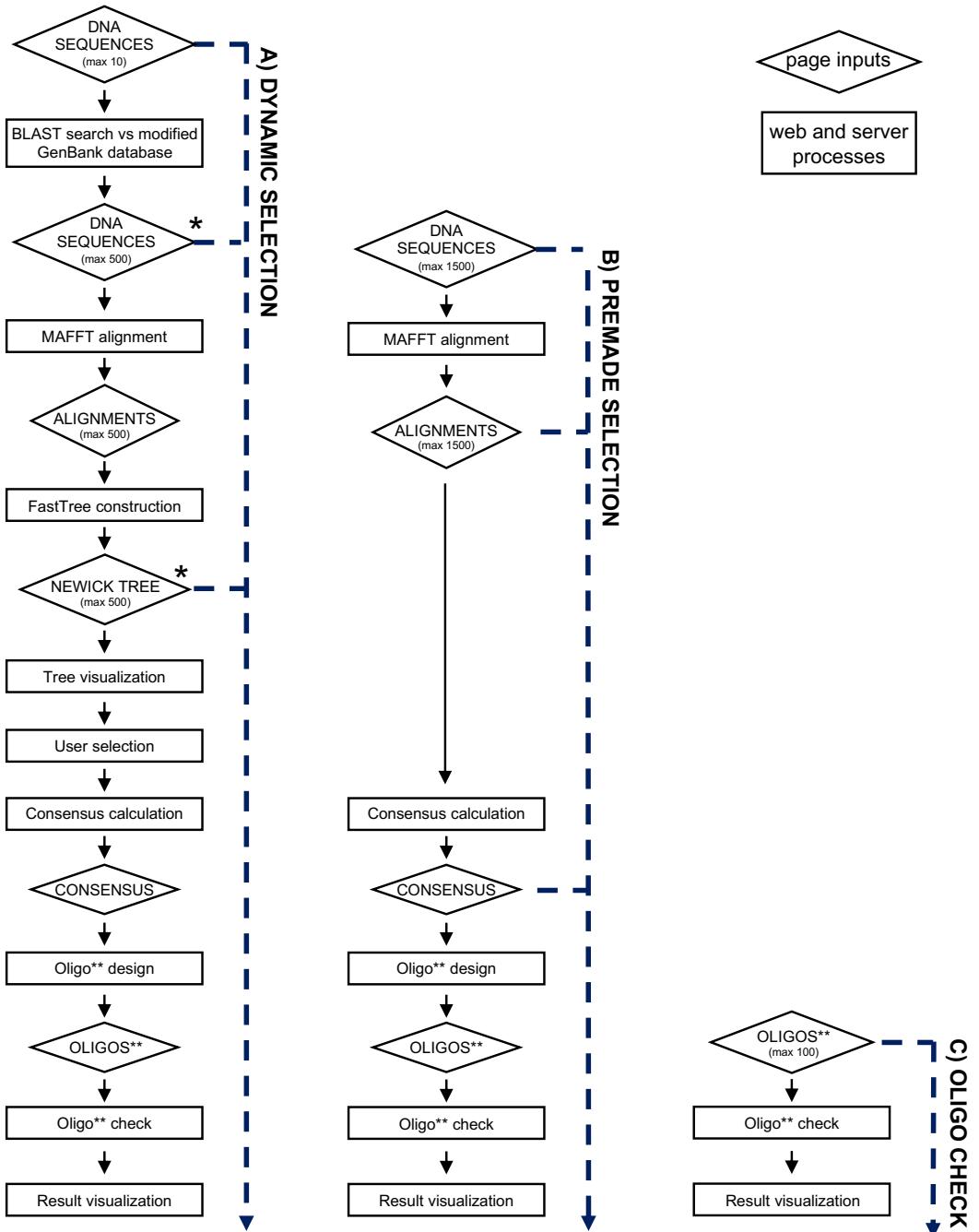


Figure 2. Scheme of the three different input page possible inputs and workflows: a) Dynamic selection, b) Premade selection and c) Oligo check. All the inputs are in the rhomboid boxes whereas the server-side processes are in the rectangles. Through the Dynamics selection page the user can input three different kind of data: up to 5 genes or DNA regions of interest, up to 500 DNA sequences and a newick tree (together with an alignment file). The Premade selection page permits the uploading of up to 500 DNA sequences, DNA alignments or directly the consensus sequence that will be used for the oligo design. In the Oligo Check page only the uploading of predesigned oligos is allowed. Different processes on the server-side of PhyloPrimer will start in relation to which data was uploaded.

and the retrieved GenBank entries. In case more than one sequence is uploaded, the DNA sequences must be highly similar, for example, they can represent the same gene from different organisms, otherwise the tree will become too populated making the visualization difficult. The DNA sequences also need to represent the same DNA fragment. For instance, if the desired amplicon is a full-length gene, the user must upload a full-length gene; if the user wants to amplify a specific gene region he can either upload the entire gene or that specific region. In other words, the DNA sequence the user uploaded must contain the region to amplify and if more than one gene was uploaded, they must all represent the same region. PhyloPrimer also imposes a length limit of 6000 bp to each sequence in order to avoid the upload of entire genomes and to encourage the upload of only the DNA segment of interest. All these limitations are necessary for the successful construction of the consensus that will be used for the oligo design, more details on this will be given in Section 4. The genes can be uploaded either with a file or pasting the sequences inside the text area input field but these two options are mutually exclusive therefore if a file was uploaded, the text area must be empty (and vice versa).

- Up to 500 DNA sequence. The uploaded DNA sequences must be at least 4 as PhyloPrimer will directly run a MAFFT alignment on the uploads and the software will not work with less than 4 sequences. PhyloPrimer will then visualize the guide tree for the dynamic visualization. In addition to the fasta file, we ask for two tab-delimited files: the first file reporting the sequence names present in the fasta file in the first column and the taxid in the second column; the second file reporting the sequence names in the first column and the gene/protein information for that entry in the second. These files are optional, in case not uploaded, none of this information will be displayed in the tree.
- A phylogenetic tree in newick format. The maximum number of allowed tree leaves is 500. PhyloPrimer will directly print the uploaded tree for the dynamic visualization. In addition to the newick file, we ask for the file reporting the DNA alignment that was used for the newick construction and for two tab-delimited files: the first file reporting the sequence names present in the fasta file in the first column and the taxid in the second column; the second file reporting the sequence names in the first column and the gene/protein information for that entry in the second. These last two files are optional, in case not uploaded, none of this information will be displayed in the tree.

Once the inputs have been uploaded, the user can click on the button **Check Uploads** and the section **Uploads** will appear (Figure 4). If there is any error with the data, the section frame will be red and there will be an error message stating the problem. Some of the most common problems are the wrong input format, errors in the file name and non uniqueness in the sequence headers (more details in Section 3). Otherwise, a blue section frame indicates that the inputs are correct and the user gets prompted with the number of uploaded DNA sequences, alignments or tree leaves (depending on which kind of input was uploaded). In case the user uploaded more than one gene in the first input field, PhyloPrimer also performs a MAFFT alignment and checks if there are many gaps in the alignments using it as a proxy for checking if the same gene fragment was uploaded, if more than the 50% of any of the sequence alignment was constituted by gaps, the software will prompt a warning. Also, if genes were uploaded in the first input area, there will be a short list of BLAST parameters that can be changed by the user (Figure 4):

- identity percentage: the percentage of bases shared between the query and the subject sequence; it can range between 60 and 100%. Default: 80%.
- coverage percentage: the percentage of bases of the query sequence that are covered by the subject sequence; it can range between 60 and 100%. Default: 90%.
- evalue: the e-value which indicates the probability of finding an alignment by chance. Default: 0.01.

Once the user is happy with all the uploaded data and has set the BLAST parameters, he can click on the button **Create Phylogenetic Tree**. Depending on which kind of input was uploaded, PhyloPrimer will start different processes. If the user uploaded genes thought the first input area, PhyloPrimer launches the BLAST search integrating the parameters selected by the user (identity, coverage and evalue).

```
blastn -db modified_GenBank -query input -task megablast -max_target_seqs 2500
-pperc_identity identity -qcov_hsp_perc coverage -eval evalue -outfmt 5
-num_threads 35
```

The BLAST search is performed against a modified GenBank database where all the Metazoa and Streptophyta sequences were removed and all the genome annotations were included (more details in Section 7). Setting the correct identity and coverage is very important for the sequence retrieval. A low identity score may retrieve genes that do not belong to the gene family of interest and the risk of using a low coverage is that the gene region to amplify is excluded as BLAST only retrieves the DNA region that has been aligned with the query sequence. This will depend on the divergence level of the gene of interest. Our suggestion is to play with the BLAST parameters but keep in mind that it will take a couple of minutes each time. In case no sequences were retrieved from the BLAST search, we suggest lowering the BLAST parameters.

DYNAMIC SELECTION

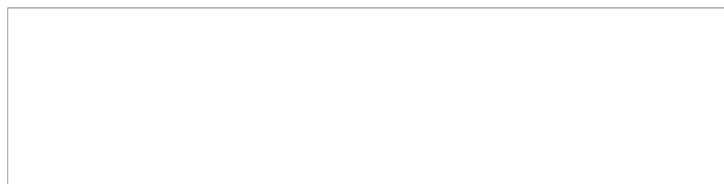
PhyloPrimer will process the data in different ways depending on your inputs (more details on this below) and will provide visualization support to choose the DNA sequences that will be used for the design of the oligos (i.e. primers and probes) in form of a dynamic tree. You have 3 different alternatives for the uploading of your data:

INPUTS

Up to 10 DNA sequences

PhyloPrimer will first perform a BLAST search against a subset of the GenBank database containing only microbial sequences, then will run a MAFFT alignment and finally will output a dynamic tree on all the input sequences and the retrieved GenBank entries. All the DNA sequences must be in [fasta format](#). The maximum number of allowed sequences is 10 and none of the sequences can be longer than 6000 bp. If you submit more than one gene make sure that they belong to the same gene family or that they have highly similar sequences (e.g. same gene belonging to different organisms). If you upload distant related DNA sequences the dynamic tree will be too populated making the result visualization difficult.

- Paste the DNA sequence(s) in the below text area. The input must be smaller than 5 Kb (the input will be trimmed in case it is bigger).



◀a

— OR —

- DNA sequence file. The maximum file size is 50 Mb.

[methyl_coenzyme_M_reductase.fasta](#)

Up to 500 DNA sequences

PhyloPrimer will directly run a MAFFT alignment on the input sequences and will print the guide tree for the dynamic visualization. All the DNA sequences must be in [fasta format](#). The maximum number of allowed DNA sequences is 500 (with a minimum of 4 sequences). In addition to the fasta file, we ask for two [tab-delimited files](#): the first file reporting in the first column the accession numbers present in the fasta file and the taxid in the second column; the second file reporting the accession numbers in the first column and the gene/protein information for that entry in the second. These files are optional, in case not uploaded, none of this information will be displayed in the tree.

- DNA sequence file. The maximum file size is 50 Mb.

No file selected.

- Tab delimited file with accession numbers and taxid information (optional). The maximum file size is 50 Mb.

No file selected.

- Tab delimited file with accession numbers and protein/gene information (optional). The maximum file size is 50 Mb.

No file selected.

◀b

Phylogenetic tree in newick format

PhyloPrimer will directly print the uploaded tree for the dynamic visualization. The tree must be in [newick format](#). The maximum number of allowed DNA sequences is 500. In addition to the newick file, we ask for the [alignment file](#) that was used for the newick construction and for two [tab-delimited files](#): the first file reporting in the first column the accession numbers present in the fasta file and the taxid in the second column; the second file reporting the accession number in the first column and the gene/protein information for that entry in the second. These last two files are optional, in case not uploaded, none of this information will be displayed in the tree.

- Newick tree file. The maximum file size is 50 Mb.

No file selected.

- DNA alignment file. The maximum file size is 50 Mb.

No file selected.

- Tab delimited file with accession numbers and taxid information (optional). The maximum file size is 50 Mb.

No file selected.

- Tab delimited file with accession numbers and protein/gene information (optional). The maximum file size is 50 Mb.

No file selected.

◀c

Figure 3. Screenshot of the Inputs section of the Dynamic selection page. This section is divided into three input areas where each of them can be enabled/disabled by clicking on its radio button. Here, for example, a) is active whereas the b) and c) are inactive.

PhyloPrimer retrieves up to 2500 matches. If more than 2500 matches are found, they will automatically be trimmed by BLAST. Once the BLAST search has terminated, PhyloPrimer groups together the sequences that

are present more than four times between the user inputs and the BLAST outputs. The newly formed groups are named with one of the sequence names that are in the cluster. The maximum number of nodes allowed in the PhyloPrimer phylogenetic tree is 500. Thus, if more than 500 unique sequences were retrieved from the BLAST search, PhyloPrimer will select only the best 500 matches for the alignment and tree construction. In case more than one gene was uploaded, per each gene the software will retrieve a maximum of unique BLAST matches corresponding to 500 divided by the number of genes. For example, if the user uploaded 5 genes, maximum 100 unique sequences per gene will be visualized on the phylogenetic tree.

If there are at least four unique sequences, the software performs a MAFFT alignment on the BLAST results together with the input sequences. A MAFFT alignment is also performed in case the DNA sequences were uploaded in the second input area.

If there are fewer than 200 sequences, the following command is run

```
mafft --localpair --maxiterate 1000 --thread 30 --quiet input > output
```

Otherwise, if there are more than 200 sequences, PhyloPrimer runs

```
mafft --maxiterate 1000 --thread 30 --quiet input > output
```

The multiple alignment is then the input for FastTreeMP which infers approximately-maximum-likelihood phylogenetic trees with a GTR+CAT model

```
FastTreeMP -nt -gtr -boot 1000 < input > output
```

The constructed tree is not accurate for phylogenetic visualization purposes but it provides a good sequence clustering to explore the sequence similarity. If the user wishes to look at more accurate trees, it is possible to directly upload a newick tree (PhyloPrimer also allows the visualization of bootstrap values).

In case the workflow involved a BLAST search, PhyloPrimer renames all the user gene sequences to avoid any clash between the user sequences and the sequence accessions retrieved from the BLAST database. In case no BLAST search was required, the software substitutes only the sequence accessions when they have 20 or more characters. PhyloPrimer renames the sequences with the label ‘USER_INPUT’ followed by a progressive numbers (e.g. USER_INPUT1, USER_INPUT2, etc).

Once PhyloPrimer has created a newick tree, the user can visualize the section Dynamic Tree. This section has a red frame and an error message in case something went wrong (e.g. no BLAST matches were found). Instead, if PhyloPrimer completed the run successfully this section will present different elements (Figure 4):

- a paragraph reporting how many BLAST sequences were found (if a BLAST search was run) and how many unique sequences were included in the tree.
- buttons **Phylum**, **Class**, **Order**, **Family**, **Genus** and **Species** which can be used for switching the taxonomy information that is reported at the right of the dynamic tree.
- tree zoom bar.
- a newick tree which can be used for the dynamic exploration of the user sequences and the GenBank sequences (in case the user uploaded the data in the first input area). The user can click on any node to select/deselect the terminal leaves which report the sequence header and, where available, the sequence taxonomy. The selected sequences will appear blue-highlighted. Details on the sequence of the BLAST search can be visualized by hovering on the nodes. The user sequence(s) are reported in the tree in bold red characters.
- a scale bar reporting the sequence divergence. e.g. a value of 0.08 means that every scale bar length the sequences have the 8% of the bases that differ.
- buttons to **Select all the nodes** and **Deselect all the nodes** can select or deselect all the tree leaves if clicked.
- a table reporting the correspondences between the PhyloPrimer name (first column) and the user original name (second column). In the tree only the name in the first column is visualized.
- if PhyloPrimer found clusters composed by more than four identical sequences, it will group their entries together. In this case, the software reports the cluster details in a table (not shown). The table reports the cluster name, the number of sequences clustered together, the sequence headers and taxonomy information at the species level.

Once the user has selected the sequences of interest, he can click on the button **Create Consensus**. This will trigger the consensus sequence calculation. If the user selected all the sequences from the tree, only one consensus (the positive consensus) is calculated. On the contrary, if only certain sequences were selected, PhyloPrimer calculates two consensus sequences: the positive consensus (from the selected sequences) and the negative consensus (from

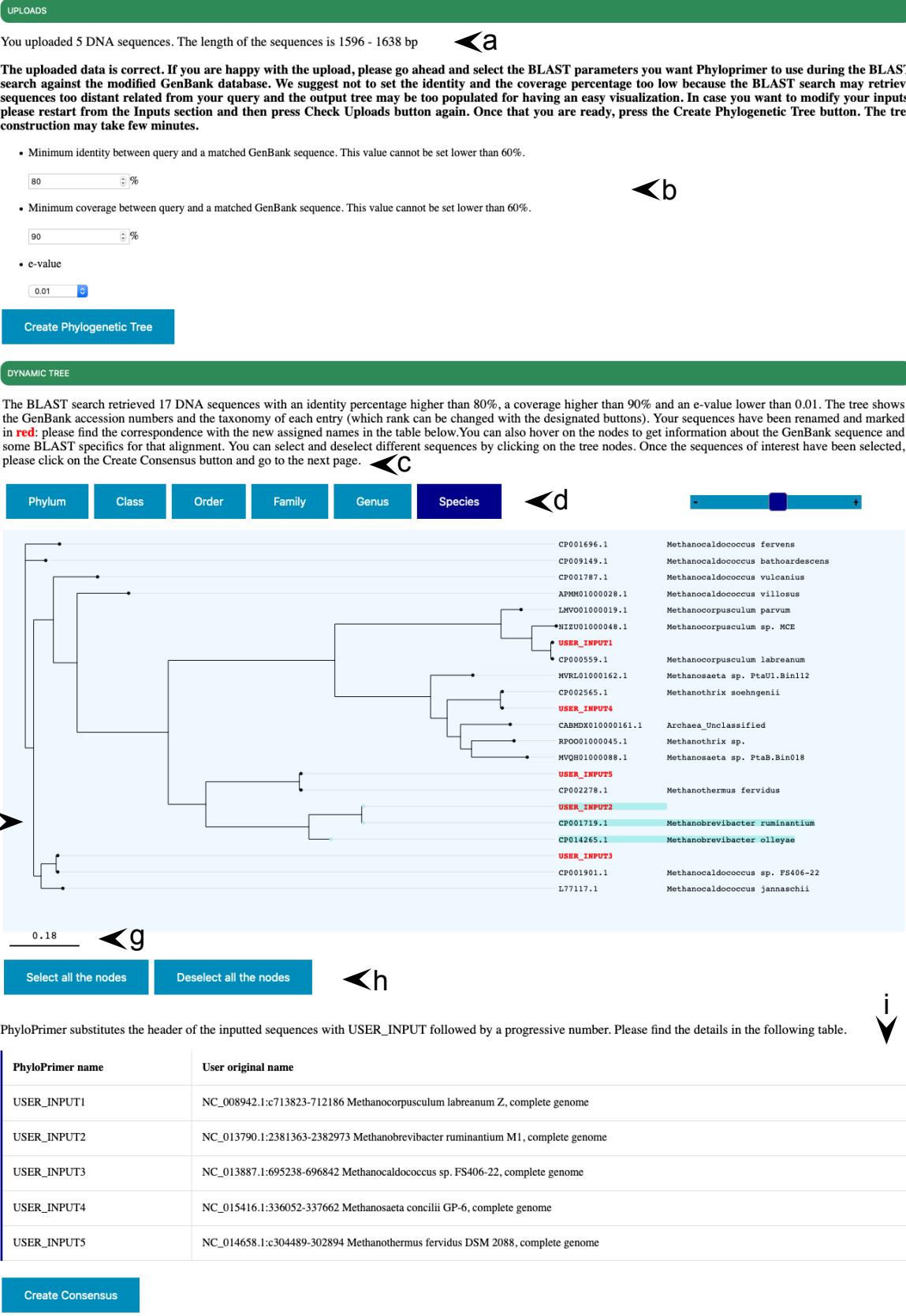


Figure 4. Screenshot of the section Uploads and Dynamic Tree of the Dynamic Visualization page where the different page elements are presented as a) paragraph reporting information on user sequences, b) BLAST parameters, c) paragraph reporting information on BLAST matches, d) taxonomy buttons, e) zoom bar, f) phylogenetic tree, g) scale bar, h) node select/deselect buttons, i) user sequence table.

the sequences that were not selected; Figure 5 and 7). The positive consensus is then used for the oligo design

CONSENSUS

PhyloPrimer calculated two consensus sequences: one out of the selected sequences (positive consensus) and the other one out of the sequences that were not selected (negative consensus). The oligos will be designed on the consensus sequences and by default PhyloPrimer will report in the Result visualization mainly the oligos that maximise the difference between the two consensus sequences.

The 15% of the bases in the consensus sequence are represented by degenerated bases.

If you are happy about the result, please click on the Oligo Design button and go to the next page.

Oligo Design

Figure 5. Screenshot of the section Consensus of the Dynamic Visualization page where only certain sequences were selected from the tree and therefore both positive and negative consensus sequences were outputted.

and the negative consensus only to look at the differing bases between the targeted and non-targeted sequences. The consensus calculation may take few seconds and when finished the sequences will be printed in a new section called Consensus. PhyloPrimer also calculates how many degenerate bases were created in the positive consensus and will suggest the maximum number of degenerate bases to include in the oligo sequence (this is an oligo design parameter that can be set in the oligo pages): it will suggest to allow 3 degenerate bases in the oligo sequence if more than the 20% of the consensus is constituted by degenerate bases, 2 bases if the degenerate bases are more than 10%, otherwise PhyloPrimer will suggest 1. If more than the 20% of the bases in the positive consensus were degenerate bases, PhyloPrimer will report a warning saying that the user may have to design several oligo assays in order to target all the selected sequences (selecting fewer sequences from the different tree clusters). This happens when the selected sequences do not have a conserved region (more details in Section 4). If re-running the same analysis with the same DNA sequences, the user should keep in mind that the order in which the DNA sequences are uploaded influences the alignment and the tree formation. Therefore, in order to be sure to get always the same cluster's order in the tree, the sequences must be uploaded in the same order.

Once the user is satisfied with the consensus sequence, the **Oligo Design** button can be clicked and he will redirect to the oligo pages.

2.2.2 Premade selection

The **Premade selection** page is used when the user has already selected the sequences he wants PhyloPrimer to use for the consensus sequence calculation and the oligo design. Therefore no dynamic tree visualization is available through this page as there is no need to explore closely related sequences. As for the Dynamic selection, when this page is first loaded, only the Inputs section is visualized. The setup of this section is similar to the one we saw before: there are three radio buttons and clicking on any of those enables/disables different input fields. The three possible inputs are either DNA sequences, DNA alignments or a consensus sequence:

- Up to 1500 DNA sequences. PhyloPrimer will first align all the DNA sequences with MAFFT (see the command reported in the previous section) and then calculate a consensus sequence that will be used for the oligo design. In case only one DNA sequence is uploaded, it will be treated as a consensus sequence. All the DNA sequences must be in fasta format.
 - Up to 1500 DNA alignments. PhyloPrimer will calculate a consensus sequence that will be used for the oligo design. In case only one DNA alignment is uploaded, it will be treated as a consensus sequence. All the DNA sequences must be in fasta format.
 - One DNA consensus sequence. PhyloPrimer will directly use the consensus sequence to design the oligos.

Only one sequence must be uploaded and it must be in fasta format.

Once the data are uploaded, the user can click on **Check Uploads** to trigger PhyloPrimer's data checks. At the end of the check, the Uploads section appears. If something is wrong the section frame will be red and an error message will help to troubleshoot the problem. If all the data are correct, the frame will be blue and there will be a short paragraph saying how many sequences were uploaded. If the user is happy with the uploads, he can then click on the next button which would be either **Check Consensus** in case the user uploaded the consensus or **Create Consensus** in case PhyloPrimer has to create the consensus from multiple DNA sequences or alignments. The first button will trigger PhyloPrimer's checks on the consensus sequence and the second one will trigger the consensus calculation. In any case PhyloPrimer will output the positive consensus in the Consensus section and clicking the **Oligo Design** button will lead the user to the oligo design pages.

2.2.3 Oligo check

The **Check Uploads** page only allows the input of oligo sequences that the user wants to be checked from PhyloPrimer. The oligos will be checked for secondary structures, melting temperature and in silico targeted organisms. These are the same checks that the oligos designed by PhyloPrimer undergo. The oligos can be uploaded as primer pairs, primer pair/probe assays and single oligos. The input must be consistent so either all the oligos need to be primer pairs, or primer pair/probe assays or single oligos and they must be maximum 100. This page presents different sections when opened:

- Inputs where the oligo sequences can be uploaded.
- BLAST Check where the user can upload a fasta file in case he wants PhyloPrimer to perform a BLAST search against specific sequences.
- PCR Conditions where the user can set the PCR conditions PhyloPrimer should use for the melting temperature and secondary structure calculation.
- Result Data where the only mandatory field is the email address as all the results are sent by email.

More information about the last three section can be found at the Section 2.3.1.

2.3 Oligo pages

At this point all the user inputs have already been uploaded and the positive consensus (and negative if needed) has already been constructed by PhyloPrimer. There are three different pages that can be selected by the navigation bar at the top of the page and they all report preset parameters for the oligo design but they can all be changed by the user. The different pages are for the **Primer Design**, **Primer and Probe Design** and **Single Oligo Design**. As in the input pages, any time the user changes an input fields, the input box will turn blue. Once all the necessary parameters have been customized by the user, he can click on **Check Parameter** button and all the parameters will be checked and printed in the new section Oligo Parameters. If there is any error in the data, the section frame is red and the wrong parameters are reported with bold red characters, otherwise the frame will be blue and the user will be able to click on the **Oligo Design** button. Once clicked, PhyloPrimer will show the submission page and will start the oligo design and check on the server side.

If the Dynamic Selection modality was used, the user is also given the possibility to go back to the initial tree and select more clusters from the same exact tree. In this way it is possible to construct multiple oligos, one (or more) for each tree cluster. This approach can also be used to try out different oligo parameters for the same cluster and sequences. Especially when a gene is very diversified, it may indeed be necessary to design multiple primers to tackle all the diversity.

This page uses GDPR cookies to memorize the user preferences and speed up the parameter selection. In the following sections are the details for each of the oligo pages.

2.3.1 Primer Design

This page is divided in the following sections: Primers, Primer Pair, Area Selection, BLAST Check, Secondary Structure, PCR Conditions, Visualization Criteria and Result Data.

User data

- Project ID: project name. The project name will be included in the email object and in the Result Page. When launching more than one job through the PhyloPrimer interface it is essential to distinguish between the different PhyloPrimer results.

- Email: all the results will be received by email. The email address, together with the oligo design results, will be kept for one month and then deleted from our server and will not be shared with any third parties. In case the user wants to be included in the PhyloPrimer mailing list, he can click the check box ‘I want to subscribe to the mailing list’. This mailing list will be used exclusively to receive updates on the software. The user can unsubscribe from the mailing list at any point by emailing gilda.varliero@bristol.ac.uk.

Primers

- Primer length: length of each primer [positive integer]. Default minimum: 18 bases. Default maximum: 24 bases.
- Melting temperature (T_m): melting temperature for each primer [positive number]. Default minimum: 54 °C. Default maximum: 64 °C.
- Homopolymer length: maximum number of same-base repeat [positive integer higher than 2]. Default maximum: 3 bases. e.g. AAA, CCC, GGG and TTT.
- Dinucleotide repeat length: maximum number of bases part of a dinucleotide repeat [even positive integer higher than 3]. Default maximum: 6 bases. e.g. ATATAT, GCGCGC, GAGAGA.
- GC clamp: number of Gs and Cs in the last 5 bases of the 3' end [positive integer]. Default minimum: 2 bases. Default maximum: 4 bases.
- GC content: percentage of Gs and Cs in each primer [positive integer between 0 and 5]. Default minimum: 40%. Default maximum: 60%.
- Number of allowed degenerate bases [positive integer]. The default number varies in relation to the degenerate base content in the oligo sequence. The default value is set to 1 if the generate bases are less than the 5% of the oligo sequence, to 2 if they are less than 20% and to 3 If this value is higher than 20%. N.B. This option does not impose the addition of degenerate bases to the primer sequences but it allows the insertion on degenerate bases only if present in the consensus sequence and if necessary. In case this value is not set to 0:
 - Substitutions: degenerate bases allowed inside the primers. Options: 2-base substitutions (R, Y, S, W, K and M), 3-base substitution (B, D, H and V) and 4-base substitutions (N). Default: 2-base substitutions and 3-base substitution.
 - Avoid degenerate bases at the 5' end for the first [positive integer]. Default: 5.
 - Avoid degenerate bases at the 3' end for the first [positive integer]. Default: 2.

Primer Pair

- Amplicon length: length of the target DNA region [positive integer]. Default minimum: 200 bases. Default maximum: 600 bases.
- Melting temperature difference (T_m) between forward and reverse primers [positive number]. Default maximum: 5 °C.
- Annealing temperature (T_a): annealing temperature [positive number]. Default minimum: 50 °C. Default maximum: 60 °C.

Area Selection

PhyloPrimer reports the positive consensus sequence that will be used for the oligo design (Figure 6a). In case only the positive consensus exists, all the consensus bases will be black. If also a negative consensus was calculated, the bases will have different colors in relation to how the bases of the two consensuses differ at each position. All the gaps are removed from the positive consensus. The two bases surrounding an area where a gap region was present in the positive consensus but not in the negative are marked with bold characters. Vice-versa, any positive consensus base corresponding to a gap on the negative consensus is colored with blue. Where, for a certain position, the base between the two consensuses differs, that base is reported in red (Figure 6b). The user can highlight on the consensus sequence only one area per oligo. The area represents the DNA portion where a specific oligo (e.g. forward primer) must be designed and PhyloPrimer will search for a specific oligo only in the indicated area. This will shorten the analyses but may lead to much less oligo choice as it will reduce the area that the software will span for the research of optimal oligos. For highlighting an area the user must first select the area on the consensus and then click the appropriate button:

- **Highlight the selected area for the forward primer design:** if clicked it will highlight the selected area on the consensus sequence. During the oligo design, the forward primer will be designed uniquely in this area.

- **Highlight the selected area for the reverse primer design:** if clicked it will highlight the selected area on the consensus sequence. During the oligo design, the reverse primer will be designed uniquely in this area.
- **Cancel all the selections:** if clicked this button will removed all the selections from the consensus sequence.

BLAST Check

At the end of the oligo design PhyloPrimer will run a BLAST search to determine which GenBank sequences are targeted by the oligos and whether there are aspecific targets. In this section, the user can upload an additional fasta file whose sequences will be used by PhyloPrimer to perform an extra BLAST search.

Secondary Structure

- Hairpin ΔG [number]. Default: -5 kcal mol⁻¹.
- Self dimer ΔG [number]. Default: -3 kcal mol⁻¹.
- Cross dimer ΔG [number]. Default: -5 kcal mol⁻¹.
- Temperature for secondary structure calculation. Secondary structure ΔG are usually calculated at a temperature of 25 °C or 37 °C. However, when considering the secondary structures in PCR reactions, the temperature should ideally be set to the PCR elongation temperature [positive number]. Default: 72 °C.

PCR Conditions

- Monovalent concentration [positive number]. The most commonly used ions in PCR reactions are Na⁺, K⁺ and Tris⁺. Default: 5 mM.
- Mg²⁺ concentration [positive number]. Default: 1.5 mM.
- Oligo concentration [positive number]. The concentration of all the oligos in the PCR reaction. Default: 2 uM.
- dNTP concentration [positive number]. Default: 2 mM.

Visualization Criteria

In this section the user can select what PhyloPrimer should take in consideration in the scoring system.

- PhyloPrimer will assign scoring points to primer pairs that have a T_m difference between the forward and the reverse primer lower than 1 °C. Default: true.
- PhyloPrimer will assign scoring points to oligos that have ΔG values higher than -1 kcal mol⁻¹. Default: true.
- PhyloPrimer will assign scoring points to oligos that do not have degenerate bases in their sequences. Default: true.

If only certain sequences were selected from the tree and PhyloPrimer created both a positive and a negative consensus:

- PhyloPrimer will assign scoring points if differing bases (between positive and negative consensus) fall into the oligo sequence at the last two bases of the 3' end. Default: true.
- PhyloPrimer will assign scoring points if differing bases (between positive and negative consensus) fall into the primer sequence. Default: true.

If the tree selection was used:

- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to species that were selected from the dynamic tree. Default: true.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same genera of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same families of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same orders of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same genera of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same classes of the selected sequences. Default: false.

a)

```

1- ATGATRYTATTGGARGTAAAAATGTCWCWAAAATTTGGAGATAARGW -50
51- AGTTTAAAAACATATCMTTYACATRRADGAAGGAGAGWCATTAGGA -100
101- TWTTRGGAAARAGTGGAGCWGAAAATCWGKYTRTRCATATGTTAAGR -150
151- GGAATGGATGGTATGARCCWACTGARGRCARATAATTAYCAYGTCTC -200
201- TTAYTGTGAAAAMTGYGGYATGTWGATGTYCCTCAAARGCWGGAAVYC -250
251- CATGTAaaaATGTGGARRWGARCTWAAAATAGAAAGTDGATTTTGG -300
301- AATGAYAAAAAATACACCTAYAAYTTAAAAGAAAATTGCTATAATGCT -350
351- TCARAGAACYTTYGCTTATATGGRGARAAWCWGTCTTGAAAYATCT -400
401- TAGARGCTTACATCARGCWGGYTATGAAGGAAARGCWATTGATWTR -450
451- GCAKTAARTTAATWAAAATGGTTAAGTTRGAGCATAGAATAACMCACAT -500
501- WGCragagatytaAGTGGAGGRGAAAGCAGAGRGTWGTTTAGCAAGRC -550
551- AAATWGCTAAAGARCCRRTTATATTYTTAGCTGATGARCCWACTGGRACT -600
601- TTAGAYCCTCAAACWGCWAAAYTRGTTCAYTCWGCYTTAAAAGAMCTTGT -650
651- YATWAARAATAAGATAAGYTTAATTAAACATCTCACTGCCAGAGGTTA -700
701- TTGCTGARYTAACDGRAARGCWATYTGGTTRGATAARGGGAGAAATTATA -750
751- ATGGAAGGWACTTCWGAGGARRTGTTAAYAAATTYATGAAACAGTWAA -800
801- AGARTTTAARAAACCWGAARYAGRDTW GARATTAAAGARGAYATTATAA -850
851- RRTTAGAAAATRHTCAAAR CAYTACTGTTCTGTTGARAGAGGRRTWRTH -900
901- AAAGCAGTTGATRRWGTWASYTTAAAYATTAGAGAGAWRGAAATMTTYGG -950
951- TTTAGTWGGARCWAGTGGDGCTGGAAAAACWACATTAGCAAAGATTATHG -1000
1001- CYGSAGTWYTWC CWCCWTCAAAGGWAARTATTGGTTAGAGTWGGAGAT -1050
1051- GAATGGGTWGA TATGACYMAACCHGMCCWAYRGGWAGAGGAAGRGTAA -1100
1101- GAGRTATATTGGWATAYTRTTCCAAGAATATGCYYTCTATCCACAYAGAA -1150
1151- CWATAYTRSAAAAYTTAACAGAGGCTATWGTTTAGARCTTCCAGATGAA -1200
1201- TTTGCAAGRATGAARGCRRTTYATACRYTARYKTCWGTWGGWTTYAGTGA -1250
1251- AGAAGARGCAGARGAAATT TAGATAARTATCCTCATGARTTRAGTGTGTTG -1300
1301- GAGARAGRCATAGRTGCTTAGCMCARGTTTAATAAAAGAGCCAAGR -1350
1351- GTTGTATMTTAGATGARCCWACWGGRACAATGGAYCCAATACAAGAAA -1400
1401- TAYHGTGCTGAATCAATHCATAAATCAAGRRYWGAGYTRGARCAAACAT -1450
1451- ATATTATTGTTCACAYGAYATGGAYTTGTWTTRAATGTYGTGATAGA -1500
1501- GCWGGATTGATGAGAAAYGGWAARTTAATWAAAGTTGGTAARCCAGAGGA -1550
1551- RATWGTGCBTTATTAACWGAGGAGGAGACARGAGATGTTGGRCAGA -1600
1601- AGT -1603

```

b)

| | |
|-----------------------------|---|
| visualized consensus | ARGCWGGAAVYCCATGTAAAATGTGGARRWGARCTWAAAATAGAA GTDGATTTTGGAAATGAYAAAAAATACACCTAYAA |
| +consensus | ↑ |
| -consensus | ARGCWGGAAVYCCATGAAAAATGTGGARRWGARCTWAAAAAATAGAA ATGMRGGARRAMMMTGY---VWVTGYGGWGRHHWYTHMRBCMRWWSAH |
| +consensus | GTDGATTTTGGAAAT---GAYAAAAAATACAC-----CTAYAA |
| -consensus | RYHGAYTTYKKRDMKSYRGRWAKAWMGADVMKSTSARRSSH RM |

Figure 6. PhyloPrimer consensus representation. a) reports how the consensus sequence is visualized in case PhyloPrimer created both a positive and a negative consensus. b) reports how the highlighted section on a) was created. PhyloPrimer reports only the positive consensus and the marked letters indicate the differences between the two sequences: red letters indicate that at that positions the two sequences presented differing bases, blue letters indicate positions where there are bases on the positive consensus but gaps in the negative and bold letters flank regions where there were gaps on the positive consensus but bases on the negative consensus. Please note that a degenerate base is marked as differing only if that base does not contain the correspondent base of the negative consensus.

- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same domains

of the selected sequences. Default: false.

2.3.2 Primer and Probe Design

This page is divided in the following sections: Primers, Probe, Primers and Probe, Area Selection, BLAST Check, Secondary Structure, PCR Conditions, Visualization Criteria and Result Data.

User data

See details in Section 2.3.1.

Primers

See details in Section 2.3.1.

Probe

- Probe length: length of the probe [positive integer]. Default minimum: 18 bases. Default maximum: 24 bases.
- Melting temperature (T_m): melting temperature for the probe [positive number]. Default minimum: 54 °C. Default minimum: 64 °C.
- Homopolymer length: maximum number of same-base repeat [positive integer higher than 2]. Default maximum: 3 bases. e.g. AAA, CCC, GGG and TTT.
- Dinucleotide repeat length: maximum number of bases part of a dinucleotide repeat [even positive integer higher than 3]. Default maximum: 6 bases. e.g. ATATAT, GCGCGC, GAGAGA.
- GC clamp: number of Gs and Cs in the last 5 bases of the 3' end [positive integer between 0 and 5]. Default minimum: 2 bases. Default maximum: 4 bases.
- GC content: percentage of Gs and Cs in the probe [positive integer]. Default minimum: 40%. Default maximum: 60%.
- DNA strand for the binding of the probe. Options: antisense strand (sense probe), sense strand (antisense probe) or either sense or antisense strand. Default: antisense strand.
- Number of allowed degenerate bases [positive integer]. Default maximum: 1 base. If this value is higher than 0:
 - Substitutions: degenerate bases allowed inside the probe. Options: 2-base substitutions (R, Y, S, W, K and M), 3-base substitution (B, D, H and V) and 4-base substitutions (N). Default: 2-base substitutions and 3-base substitution.
 - Avoid degenerate bases at the 5' end for the first [positive integer]. Default: 5.
 - Avoid degenerate bases at the 3' end for the first [positive integer]. Default: 2.

Primer pair and Probe

- Amplicon length: number of bases the target amplicon should have [positive integer]. Default: 200 - 600 bases
- Melting temperature (T_m) difference between forward and reverse primers [positive number]. Default maximum: 5 °C.
- Melting temperature (T_m) difference between primers and probe [positive number]. The probe melting temperature should always be higher than the primer one. Default minimum: 5 °C.
- Annealing temperature (T_a): annealing temperature range [positive number]. Default minimum: 50 °C. Default maximum: 60 °C.

Area Selection

PhyloPrimer reports the positive consensus sequence that will be used for the oligo design. See more details on the Area Selection in Section 2.3.1. The buttons that must be used for the area selection are:

- **Highlight the selected area for the forward primer design:** if clicked it will highlight the selected area on the consensus sequence. During the oligo design, the forward primer will be designed uniquely in this area.

- **Highlight the selected area for the reverse primer design:** if clicked it will highlight the selected area on the consensus sequence. During the oligo design, the reverse primer will be designed uniquely in this area.
- **Highlight the selected area for the probe design:** if clicked it will highlight the selected area on the consensus sequence. During the oligo design, the probe will be designed uniquely in this area.
- **Cancel all the selections:** if clicked this button will removed all the selections from the consensus sequence.

BLAST Check

See details in Section 2.3.1.

Secondary structure

See details in Section 2.3.1.

PCR conditions

See details in Section 2.3.1.

Visualization Criteria

See details in Section 2.3.1.

2.3.3 Single Oligo Design

This page is divided in the following sections: Oligos, Area Selection, BLAST Check, Secondary Structure, PCR Conditions, Visualization Criteria and Result Data.

User data

See details in Section 2.3.1.

Oligos

- Oligo length: length of the oligo [positive integer]. Default minimum: 18 bases. Default maximum: 24 bases.
- Melting temperature (T_m): melting temperature for the oligo [positive number]. Default minimum: 54 °C. Default minimum: 64 °C.
- Homopolymer length: maximum number of same-base repeat [positive integer higher than 2]. Default: 3 bases. e.g. AAA, CCC, GGG and TTT.
- Dinucleotide repeat length: maximum number of bases part of a dinucleotide repeat [even positive integer than 3]. Default: 6 bases. e.g. ATATAT, GCGCGC, GAGAGA.
- GC clamp: number of Gs and Cs in the last 5 bases of the 3' end [positive integer between 2 and 5]. Default minimum: 2 bases. Default maximum: 4 bases.
- GC content: percentage of Gs and Cs in the oligo [positive integer]. Default minimum: 40%. Default maximum: 60%.
- DNA strand for the binding of the probe. Options: antisense strand (e.g. forward primer or sense probe), sense strand (e.g. reverse primer or antisense probe) or either sense or antisense oligo. Default: antisense strand.
- Type of oligo. Options: probe-like or primer-like. Default: probe-like.
- Number of allowed degenerate bases. Default maximum: 1 base. If this value is higher than 0:
 - Substitutions: degenerate bases allowed inside the probe. Options: 2-base substitutions (R, Y, S, W, K and M), 3-base substitution (B, D, H and V) and 4-base substitutions (N). Default: 2-base substitutions and 3-base substitution.
 - Avoid degenerate bases at the 5' end for the first. Default: 5.
 - Avoid degenerate bases at the 3' end for the first. Default: 2.

Area Selection

PhyloPrimer reports the positive consensus sequence that will be used for the oligo design. See more details on the Area Selection in Section 2.3.1. The buttons that must be used for the area selection are:

- **Highlight the selected area for the oligo design:** if clicked it will highlight the selected area on the consensus sequence. During the oligo design, the oligo will be designed uniquely in this area.

- **Cancel all the selections:** if clicked this button will removed all the selections from the consensus sequence.

BLAST Check

See details in Section 2.3.1.

Secondary structure

See details in Section 2.3.1.

PCR conditions

See details in Section 2.3.1.

Visualization Criteria

In this section the user can select what PhyloPrimer should take in consideration in the scoring system.

- PhyloPrimer will assign scoring points to oligos that have ΔG values higher than -1 kcal mol⁻¹. Default: true.
- PhyloPrimer will assign scoring points to oligos that do not have degenerate bases in their sequences. Default: true.

If only certain sequences were selected from the tree and PhyloPrimer created both a positive and a negative consensus:

- PhyloPrimer will assign scoring points if differing bases (between positive and negative consensus) fall into the oligo sequence at the last two bases of the the 3' end. Default: true.
- PhyloPrimer will assign scoring points if differing bases (between positive and negative consensus) fall into the primer sequence. Default: true.

If the tree selection was used:

- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to species that were selected from the dynamic tree. Default: true.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same genera of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same families of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same orders of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same genera of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same classes of the selected sequences. Default: false.
- PhyloPrimer will assign scoring points if the oligos align to DNA sequences belonging to the same domains of the selected sequences. Default: false.

2.4 Result page

If PhyloPrimer found at least one oligo assay that reflected all the prerequisites, the user will receive a link to the final page by email. The Result Page is divided into two main sections: on the left of the page there is the **Oligo List** which reports the 100 oligo assays considered the best by the PhyloPrimer scoring system. The assays with the highest scores are first reported in the list. The oligo assay can be either constituted by a primer pair, a primer pair plus a probe or a single oligo, depending on what was requested by the user. If the Oligo List shows oligos that were designed by PhyloPrimer all the oligo sequences will be preceded by F, R, P or O if the oligo is a forward primer, reverse primer, probe or a single oligo, respectively, then by a number which is the oligo position and, separated by a line, the oligo length (e.g. F57-18-TAGACGGGCTGACGTATG : this is a forward primer which 5' end is at the 57 bases on the consensus and it is 18 bases long). In the result page all the reported positions are intended as the positions at the 5' oligo end. If the oligos were only checked and not designed by PhyloPrimer, they will be reported with a sequence name only if it was provided by the user. The **Oligo List** also presents a search window that can be use for highlighting the oligos that reflect different criteria: melting temperature between a certain range, oligo length and taxonomy correspondent to the GenBank sequences that matched against all the oligos of an oligo assay. The taxonomy search works only at genus and species level.

On the right side of the page there are all the information regarding the oligos that have been selected from the **Oligo List**. The information is divided into different pages: **Specifics**, ΔG , **GenBank BLAST**, **Your BLAST** (if the user uploaded a fasta file for an additional BLAST check), and **Archive**.

2.4.1 Specifics

This section reports the specifics for the selected oligos. First, it reports the consensus sequence with the oligo positions (not present in the Check mode). In case both a positive and a negative consensus were calculated, it reports also the differing bases between the two sequences so that the user can visualize whether the oligos fall inside a differing region (more details in Section 2.3.1 and Figure 6). On the consensus the oligo positions are also reported. Below the consensus sequence there is a table reporting:

- Oligo type: F, R, P for forward primer, reverse primer and probe, respectively (not present for single oligos).
- Sequence: the oligo sequence.
- Position: the position at the 5' end of the oligo.
- Strand: the sense of the DNA filament the oligo anneals to (not present for primer pairs).
- Length: the oligo length.
- GC%: the percentage of Gs and Cs in the oligo sequence.
- T_m : the melting temperature in °C.
- Self Dimer: the lowest ΔG in kcal mol⁻¹ measured for the self dimer secondary structures of the oligo.
- Hairpin: the lowest ΔG in kcal mol⁻¹ for the hairpin secondary structures of the oligo.
- T_a : the annealing temperature in °C. This is just indicative as this temperature can change a lot in relation to the PCR polymerase that is used (not present for single oligos).
- Cross Dimer: the lowest ΔG in kcal mol⁻¹ for the cross dimer secondary structures of the oligo (not present for single oligos).

2.4.2 ΔG

This section reports all the self dimer, cross dimer and hairpin formations that were found for the selected oligos. The page reports in details all the oligo structures with an associated ΔG value lower than 0 kcal mol⁻¹. More details on how the secondary structures are reported in Section 6.2 and 6.3.3.

2.4.3 GenBank BLAST

The selected oligos were blasted and global aligned with Bowtie against a modified version of the GenBank database (more details in Section 7). In this page PhyloPrimer reports the result of that BLAST search. If the result page is reporting primer assays, the pie chart will report the taxonomy of the GenBank sequences aligned to both the forward and reverse sequences, if we are visualizing primer pair/probe assays, the pie chart will show matches that were aligned to the forward primer, reverse primers and probe together. If we are dealing with single oligos, the pie chart will report all the the matches. The pie chart may be misleading if the user does not take in consideration the fact that different species are more studied than others and therefore more likely present in the GenBank database. Therefore the pie chart's aim is only to give an indicative idea about which organisms could be potentially targeted.

A BLAST match is considered valid if the oligo and the GenBank sequence matched with three or less total mismatches. Below the pie chart there is a table reporting all the BLAST matches and some of the match specifics. The matches are ordered by which oligos they aligned to and by the alignment qualities. The table presents the following fields:

- GenBank sequence: GenBank accession number.
- Oligo type: F, R or P whether the oligo was a forward primer, a reverse primer or a probe (not present if single oligos).
- External mismatches: the number of external mismatches.
- Internal mismatches: the number of internal mismatches.

- Alignment start and Alignment end. These two columns report the start and the end of the alignment. Looking at which is the biggest number between start and end the user can understand which strand the primer binds. Alignment start and end are followed by an asterisk in case the oligo was aligned to multiple positions of a GenBank sequence.
- Amplicon size: if the GenBank sequence was matched by both the forward and the reverse primer (not present for single oligos).
- Species (or Genus or Family or Order or Class or Phylum or Domain): taxonomy of the GenBank sequence.

In case no matches between your oligos and the GenBank database were found the message 'PhyloPrimer did not find any match between your fasta file and these oligos.' is visualized.

2.4.4 Your BLAST

As in the **GenBank BLAST**, this page reports a pie chart and a table with the BLAST results. The only difference is that the BLAST search was performed on a database created by the fasta file that was uploaded by the user, therefore this page will only be visible if that file was uploaded. All the BLAST matches are reported in the table whose fields are:

- Your sequence: your sequence header.
- Oligo type: F, R or P whether the oligo was a forward primer, a reverse primer or a probe (not present if single oligos).
- External mismatches: the number of external mismatches.
- Internal mismatches: the number of internal mismatches.
- Alignment start and Alignment end. These two columns report the start and the end of the alignment. Looking at which is the biggest number between start and end the user can figure out on which strand the primer binds. Sometimes the starting and ending position are followed by an asterisks in case the oligo was aligned to different positions of a GenBank sequence.
- Amplicon size: if you sequence was matched by both the forward and the reverse primer (not present in single oligos).

In case no matches between your oligos and the your database were found the message "PhyloPrimer did not find any match between your fasta file and these oligos." is returned.

2.4.5 Archive

From this page the user can download an archive with all this input files and all the files that were output during the oligo designing or checking. The files can be found in two different folders: Inputs and Results. The folder's content changes in relation to the uploaded data and the selected PhyloPrimer functionalities. In any case PhyloPrimer generates a random code of 20 characters which is attributed to each job and the first 10 characters are used in the file nomenclature, here called userID. Here's the scheme of the two folders detailed case by case:

inputs

1. Dynamic Selection

(a) Up to 10 DNA sequences

userID.fasta: DNA sequences
 userID.fasta.out: BLAST search output
 userID.fasta.out.tree: guide tree
 userID.fasta.out.alignment: DNA alignments
 userID.fasta.out.alignment.positive: selected DNA alignments from dynamic tree
 userID.fasta.out.alignment.negative: DNA alignments not selected in the dynamic tree (not present if all the sequences were selected from the tree)
 userID.consensus.positive: consensus calculated by the selected sequences
 userID.consensus.negative: consensus calculated by sequences that were not selected in the dynamic tree (not present if all the sequences were selected from the tree)
 userID.taxonomy: taxonomy of the selected tree sequences where 1 corresponds to Domain, 2 to Phylum, 3 to Class, 4 to Order, 5 to Family, 6 to Genus and 7 to Species (present only when the sequences selected from the phylogenetic tree are associated to a taxonomy description)

userID.info: oligo design parameter file

(b) Up to 500 DNA sequences

userID.fasta: DNA sequences

userID.tax.txt: tab delimited taxid file

userID.pro.txt: tab delimited gene file

userID.alignment: DNA alignments

userID.fasta.tree: guide tree

userID.alignment.positive: selected DNA alignments from dynamic tree

userID.fasta.out.alignment.negative: DNA alignments not selected in the dynamic tree (not present if all the sequences were selected from the tree)

userID.consensus.positive: consensus calculated by the selected sequences

userID.consensus.negative: consensus calculated by sequences that were not selected in the dynamic tree (not present if all the sequences were selected from the tree)

userID.taxonomy: taxonomy of the selected tree sequences where 1 corresponds to Domain, 2 to Phylum, 3 to Class, 4 to Order, 5 to Family, 6 to Genus and 7 to Species (present only when the sequences selected from the phylogenetic tree are associated to a taxonomy description)

userID.info: oligo design parameter file

(c) Phylogenetic tree in newick format

userID.tree: newick tree

userID.alignment: DNA alignments

userID.tax.txt: tab delimited taxid file

userID.pro.txt: tab delimited gene file

userID.alignment.positive: selected DNA alignments from dynamic tree

userID.fasta.out.alignment.negative: DNA alignments not selected in the dynamic tree (not present if all the sequences were selected from the tree)

userID.consensus.positive: consensus calculated by the selected sequences

userID.consensus.negative: consensus calculated by sequences that were not selected in the dynamic tree (not present if all the sequences were selected from the tree)

userID.taxonomy: taxonomy of the selected tree sequences where 1 corresponds to Domain, 2 to Phylum, 3 to Class, 4 to Order, 5 to Family, 6 to Genus and 7 to Species (present only when the sequences selected from the phylogenetic tree are associated to a taxonomy description)

userID.info: oligo design parameter file

2. Premade Selection

(a) Up to 1500 DNA sequences

userID.fasta: DNA sequences

userID.alignment: DNA alignments

userID.consensus.positive: consensus sequence

userID.info: oligo design parameter file

(b) Up to 1500 DNA alignments

userID.alignment: DNA alignments

userID.consensus.positive: consensus sequence

userID.info: oligo design parameter file

(c) One DNA consensus sequence

userID.consensus: consensus sequence

userID.consensus.positive: consensus sequence (formatted)

userID.info: oligo design parameter file

3. Oligo Check

userID.fasta: oligo DNA sequence

userID.info: oligo design parameter file

results

1. Primer Design

forwardList.txt: list of forward primers
reverseList.txt: list of reverse primers
assayList.txt: list of primer pairs
selfDimer.txt: self-dimer secondary structures associated with each oligo
crossDimer.txt: cross-dimer secondary structures associated with each oligo
hairpin.txt: hairpin secondary structures associated with each oligo
oligo_bowtie_nt.sam: Bowtie global alignment vs DB2
oligo_bowtie_user.sam: Bowtie global alignment vs the user database (only if it was performed)

2. Primer and probe Design

forwardList.txt: list of forward primers
reverseList.txt: list of reverse primers
probeList.txt: list of probes
assayList.txt: list of primer pairs
selfDimer.txt: self-dimer secondary structures associated with each oligo
crossDimer.txt: cross-dimer secondary structures associated with each oligo
hairpin.txt: hairpin secondary structures associated with each oligo
oligo_bowtie_nt.sam: Bowtie global alignment vs DB2
oligo_bowtie_user.sam: Bowtie global alignment vs the user database (only if it was performed)

3. Single Oligo Design

oligoList.txt: list of oligos
selfDimer.txt: self-dimer secondary structures associated with each oligo
hairpin.txt: hairpin secondary structures associated with each oligo
oligo_bowtie_nt.sam: Bowtie global alignment vs DB2
oligo_bowtie_user.sam: Bowtie global alignment vs the user database (only if it was performed)

4. Oligo Check oligoList.txt: list of oligos

selfDimer.txt: self-dimer secondary structures associated with each oligo
crossDimer.txt: cross-dimer secondary structures associated with each oligo (not if single oligos)
hairpin.txt: hairpin secondary structures associated with each oligo
oligo_bowtie_nt.sam: Bowtie global alignment vs DB2
oligo_bowtie_user.sam: Bowtie global alignment vs the user database (only if it was performed)

In the List.txt files PhyloPrimer reports all the information for forward (forwardList.txt), reverse (reverseList.txt), probe (probeList.txt), single oligos (oligoList.txt) also for the oligo assays (assayList.txt). For all these files the following information is reported:

- FORWARD (or REVERSE or PROBE or OLIGO): sequence of the forward primer (or reverse primer or probe or single oligo)
- POSITION: oligo position on the consensus
- LENGTH: oligo length
- GC: GC percentage
- TM: melting temperature
- SELF DIMER: lowest ΔG associated to self dimer formations
- HAIRPIN: lowest ΔG associated to hairpin formations
- SPECIES: species assigned to GenBank sequences BLASTed against all the oligos belonging to the designed or checked oligo assays (e.g. both forward and reverse primers when working with primer pairs). This will report ‘not BLASTed’ when no BLAST search was performed (only the best 250 or 500 oligo assays are taxonomy checked, more details in Section 4).

In the oligoList.txt and assayList.txt files also the following fields are reported:

- INDEX: serial number of the oligo assay
- SCORE: score assigned from PhyloPrimer scoring system

In the pairList.txt files also the following fields are reported:

- TA: annealing temperature
- CROSS DIMER: lowest ΔG associated to cross dimer formations

3 Input formats

The possible inputs to the PhyloPrimer workflow can be DNA sequences, alignments, newick trees or tab-delimited files. The input submission is achieved with the upload of files or, in the case of DNA sequences also with the pasting of the sequences inside a text field. The maximum size of any file is 50 Mb whereas the text box has a limit of 50 Kb. Depending on the input page, there are different rules for the maximum number of sequences can be uploaded or a maximum of the sequence lengths. The file name must not contain special characters. All the input formats are also reported in a specific Format Page can be visualized by clicking on any of the format (blue links).

3.1 DNA sequences

All the DNA sequences must be in fasta format which means that all the DNA sequences must have an header (PhyloPrimer tolerates the omission of the header in case the input is just one DNA sequence). The sequence header must start with the >sign followed by the sequence name which must be unique. The header can contain only AZ, az, 09, . , - and -, in case there some others, PhyloPrimer is going to substitute them with the sign _ and may change the user input sequences if needed (Section 2.2.1). The sequence can contain the nucleic bases A, T, G, C, and all the degenerated bases R, Y, S, W, K, M, B, D, H, V, N (case insensitive; Table 1). The sequence can be split over multiple lines but no spaces are allowed in the sequence. When the user wants to upload the consensus sequence, only one sequence is allowed.

e.g

```
>SEQUENCE1
ATGTCAGATTGCGTCAAATCGCATTCTACGGCAAAGGGGCATCGGAAGTCCACCACTCACAAAATACGCTCGGGCGCTCGT
CGACCTCGGGCAGAAAATCCCATCGTCGGATGCGATCCCAAAGCCGACTCCACCCGCTGATCCTGAACCGCAAAGCACAGGACAC
GGTTCTGCATCTCGGGCACAGGAAGGTTGGTGGAAAGACCTCGAAGTCGAGGACGTGCTCAAGACCGGCTACAAAGGTATCAAGT
ACAAGGCTCAGGAGATCTACATCGTCATGTCCGGCAGAGATGATGGCGCTCTATGCCGCAACAAACATGCCAAGGGCATCCTGAAA
TACGCCCACTCAGGCGCGTGCAGGCTCGGCTGATCTGTAACGAGGCCAGACGGATCGCAGGCTCGACCTCTCCGAGGCAT
CTGGCTGCCAGGCTCAATTCCAAGCTCATCCACTTGTGCCGCTGACAACATCGCCAGCAGCCGAACTCAGGAAGAGATGACCGT
GATCCAGTACCGCGCCGACTCCAAGCAGGCAGGGAAATATCGCGACTAGCCGAGAAGATCCATGCCAATTGGGCAAGGCACCA
GTTCCGACTCCGATACCATGGAGGAGCTCGAGGACATGCTGCTGATTTCGGCATCATGAAGACTGACGAGCAAATGCTTGCGA
ACTTCAGGCCAAGGAAGCGACGGTGGCGGCCGCCGATAA
>geosmine
ATGTCAGATTGCGTCAAATCGCATTCTACGGCAATACGCTAAAGGG
GGCATCGGCAAGTCCACCACCTCTCAGAATACGCTCGGGCGCTCGT
GACCTAGGGCAGAAAATCCTCATTGTCGGATGCGATCCCCTCTCAG
AAAGCCGACTCCACCCGCTGATCCTGAACCGAAAGCACAGGACAC
GTTCTGCATCTCGCagcgccgcatagacatagaatgACAGGAAGGT
CGAGGACGTGCTCAAGACCGGCTACAAAGGTATCAAGTGGGAAGGT
GTTGAGTCCGGCGGTCCGGAGCCGGCGTGGCTGGCGAAGGGTTGCA
GCCGGCCGCGCGTCATCACATCGATCAATTCTCGAGGAAACGGC
TATGACGATGCGACTACGCTCCTATGATGTAACGGCGGAAGGGTT
GATGTGGTGTGctagacatagacacacacatgacatgGCGATGCCGA
GGAGATCTACATCGCATGTCATGTCGGCAGAGATGATGGCGCTATGCAC
CGCCAACACATGCCAAGGGCATCCTGAAATATGCCACTCAGGCA
CGCGTGCAGGCTCGCGGCCGATCTGTAACGAGGCCAGACGGATC
>AC123
GGGCATCGGCAAGTCCACCACCTCCAAAATANNNNNNNNNNNNNTCGGGCAGAAAA
TCCTCATCGTGGATGCGATCCCAAAGCCGACTCCACCCGCTGATCCTGAACCGCAA
CGCAGGACACGGTCTGCATCTCGCAGCACAGGAAGGTTGGAGACCTCGAAGT
ACGTGCTCAAGGCCGGTRACAGAGGCATCAAGTGTGTGGAGTCCGGCGGTCCGGAGCC
```

Table 1. Degenerate base substitutions.

| Symbol | Included bases | | | |
|--------|----------------|---|---|---|
| A | A | - | - | - |
| C | - | C | - | - |
| G | - | - | G | - |
| T | - | - | - | T |
| W | A | - | - | T |
| S | - | C | G | - |
| M | A | C | - | - |
| K | - | - | G | T |
| R | A | - | G | - |
| Y | - | C | - | T |
| B | - | C | G | T |
| D | A | - | G | T |
| H | A | C | - | T |
| V | A | C | G | - |
| N | A | C | G | T |

```

GCTGCGCCGGCCGTGGCGTCATCACCTCGAT-----CGGCCTTAT
GACTACGTCTCCTATGATGTGCTCGCGATGTGGTGTGCGGGCTCGCGATGCCGA
GCCCAAGAGATCTACATCGTCATGTCCGGCGAGATGATGGCGCTATGCCGCAAAC
TCCTGAAATATGCCCACTCAGGCGCGTGGCTGCCAGGCTCGGCGGCTGATCTGTAACGACTC
TCGACCTCTCGAGGCGCTGGCTGCCAGGCTCAATTCCAAGCTCATCCACTTTATCGT
CAGCACGCGAGCTCAGGAAGATGACGGTGATCCAGTACGCGCCGGCAGGCCGGCGA
CGCTAGCCGAGAAGATCCATGCCAATTGGGCCAAGGCACCGTCCGACCCGAGTTAC
GACATGCTGTTGATTGGCATCATGAAGACCGACGAGCAAATGCACTTCACCGCCCA

```

3.2 DNA alignments

The DNA alignment must have the same format as DNA sequences. All the sequences must be the same length and the introduction of gaps as - or . is allowed.

3.3 Newick tree

The phylogenetic tree is created in newick format. The inclusion of bootstrap values is supported.

e.g

```
(AB568291.1:0.0000021975, ((CP027541.1:0.0032827700, (LN831039.1:0.0000021975,
(((CP009496.1:0.0000000000, CP001663.1:0.0000000000):0.0000000000,
CP000480.1:0.0000000000):0.0000000000, CP009494.1:0.0000000000):0.0000021975,
CP009495.1:0.000021975)88:0.0018904132)69:0.0018807825)
100:0.0409441482, CP012150.1:0.0397290108)100:0.0303804720, SEQUENCE1:0.0000021975);
```

3.4 Tab-delimited tables

These files are made by two columns, tab-delimited file, with the sequence name in the first column and the taxid (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>; Federhen, 2012) in the second column. Or, by two columns, tab-delimited file, with the sequence name in the first column and gene/protein in the second column. These files are optional.

e.g

```
SEQ1 562
SEQ2 548
SEQ3 562
SEQ4 548
SEQ5 562
```

SEQ6 562
SEQ7 170540
SEQ8 170540
SEQ9 170540
SEQ10 170540

3.5 Oligo assays

The oligo sequences can contain the nucleic bases A, T, G, C, and all the degenerated bases R, Y, S, W, K, M, B, D, H, V, N (case insensitive; Table 1). The oligo can be just a plain sequence or if there is the sequence name, it must precede the sequence and be separated by it by "-". The user can submit either single oligos (one oligo) or primer pairs (forward and reverse primers) or primer pairs plus the probe (forward primer, reverse primer and probe), one per line. The oligos belonging to the same primer pairs or primer pair/probe must be on the same row, must be separated by a tab or a space character and ordered as follows: forward primer, reverse primer and probe. The input must have only one kind of data so either all the oligos must be single oligos or primer pairs or primer pair/probe.

e.g.

```
1F-CACCACTGCCGAACTTGTTG
2F-AGACCGCCTTCACCAACA
PRIMER3-CGAACCACTCGAAGTCGGTGT
4REVERSE-ATCCAGATGAACCTCAAGAAC
GCGACACATCTCGAACGGCTAC
geosminF-CCGAACTTGTTGTACCAGC
GGTAACTGCGGTAGTAGTCGT
```

e.g.

```
geo1f-CACATCTCCAACGGCTACTC geo1R-AGGTAACTGCGGTAGTAGTCG
2F-CACATCTCCAACGGCTACT 2R-AGGTAACTGCGGTAGTAGTCGT
CACATCTCCAACGGCTACT GAAGGTAACTGCGGTAGTAGTC
f-TGGAACAAACCACTGCGTC r-GAAGGTAACTGCGGTAGTAGTC
```

e.g.

```
CACATCTCCAACGGCTACTC AGGTAACTGCGGTAGTAGTCG ACGACGCAGTGGTTGTTCCACC
1metF-CACATCTCCAACGGCTACT 1metR-AGGTAACTGCGGTAGTAGTCGT 1metP-ACGACGCAGTGGTTGTTCCA
TGGAAACAACCACTGCGTC AACGTTAAGGTAACTGCGGTAGTAGTCG CCTTGCTGCCGTACTCGATGA
f-GAACAAACCACTGCGTCGT r-GAAGGTAACTGCGGTAGTAGTC probe-CCTTGGTGCCGTACTCGATGA
```

4 Oligo Design and Scoring System

PhyloPrimer uses a consensus approach for the oligo design or, in other words, it designs the oligos from a consensus sequence. The consensus sequence can be uploaded to PhyloPrimer by the user through the Premade Selection page or it can be calculated by PhyloPrimer itself. The software constructs the consensus with the DNA sequences or alignments uploaded through the Premade Selection page or with the sequences that were selected by the user on the dynamic tree (more information on the consensus calculation can be found in the Section 6). In order for PhyloPrimer to find suitable oligos, the consensus must have one or more conserved regions, DNA regions that are in common among all the selected/uploaded sequences (in case of the Dynamic Selection or Premade Selection, respectively). If no conserved regions were present, the consensus sequence will be represented by long stretches of degenerate bases and the software will not be able to design any oligo from it. There can be different reasons for this: i) the sequence selection was too broad for the target gene family, ii) the selected sequences did not include only sequences from the same gene family, iii) the sequences represented different DNA regions of the same gene or iv) the studied gene family is very divergent. In general, it is more likely to have a conserved region in the consensus when working with closely related sequences, for example, when developing oligos for a specific species rather than for an entire gene family. However, when the aim is to develop oligos at gene level, the presence of a conserved gene region between different organisms highly depends on the gene sequence. It is very important to know the gene family object of the study and to check the consensus sequence that PhyloPrimer reports. In case the consensus presents a lot of degeneracy, it will be necessary to adjust the maximum number of degenerate bases

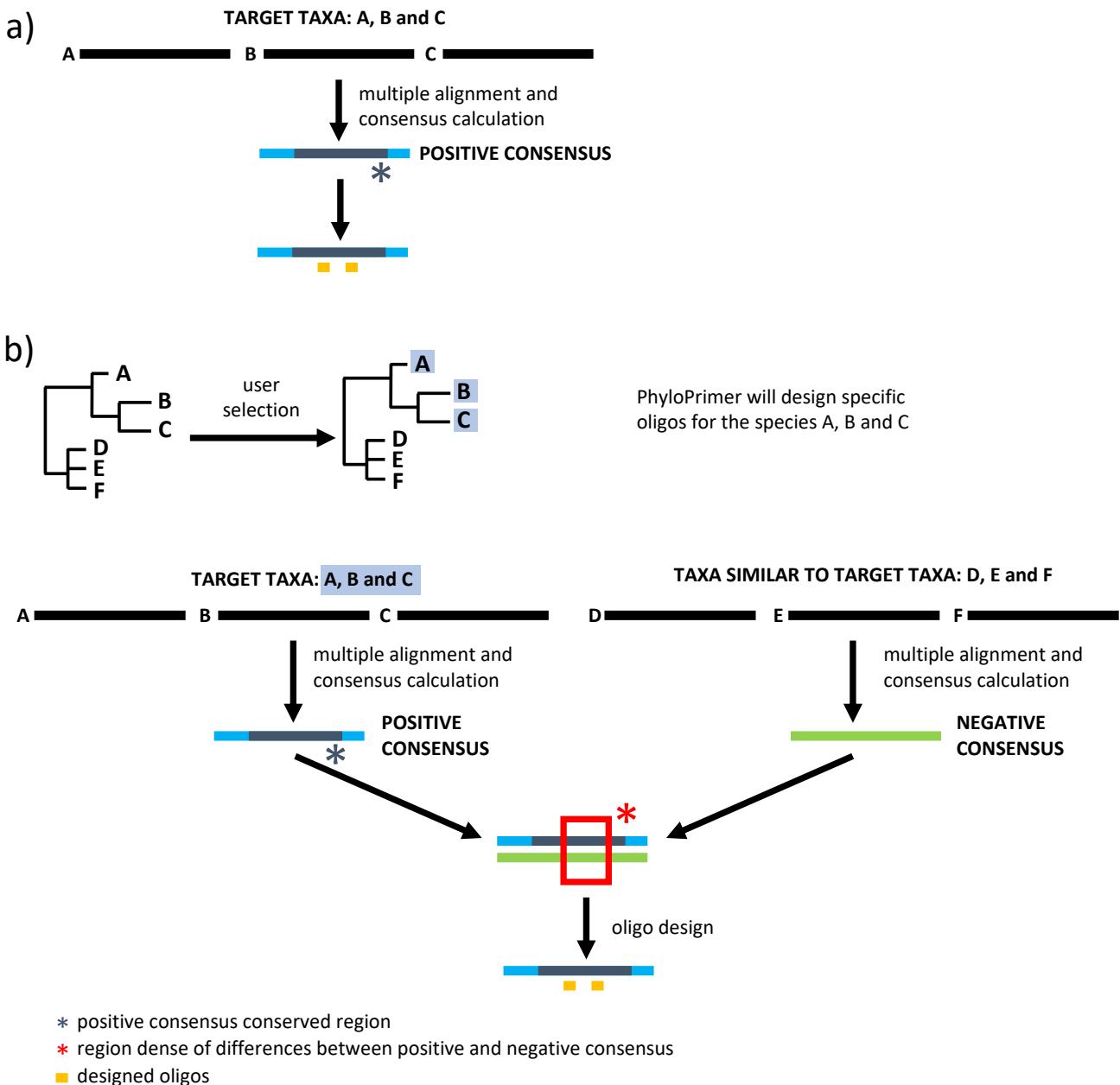


Figure 7. Positive and negative consensus calculation. The first strategy relies on the construction of only one consensus, the positive consensus. The second implies the construction of a positive and a negative consensus and it is used in the Dynamic Selection page for assuring the specificity of the designed oligos.

allowed inside the oligo sequence in the Oligo Design pages. If this does not help, the design of different oligos for different cluster of organisms should be considered.

In case the data was uploaded on the Premade Selection page, the software will find the best consensus areas and then will design the oligos from the consensus (Figure 7a). In case the data was uploaded on the Dynamic Selection page, PhyloPrimer assumes that the user wants to have taxonomic specific primers and checks the oligos for taxon specificity by default. In order to design and retrieve oligos that are specific to the sequences that were selected from the dynamic tree PhyloPrimer uses a two consensus strategy (Figure 7b). Two consensus sequences are calculated: the positive consensus is calculated by only the sequences that were selected by the dynamic tree and the negative consensus by the sequences that were not selected. All the oligos will be designed uniquely on the positive consensus. The negative will be only used to look for differing bases between the two consensus sequences. The software will use this information when scoring the oligos with the aim to retrieve the best ones to be visualized in the dynamic result page (more on this later in this section). In PhyloPrimer the conserved region of the consensus sequence is determined by the maximum number of degenerate bases that is allowed inside the oligo sequences. For instance, if the user sets the maximum degenerate base value to 1, PhyloPrimer will discard all the oligos that have more than 1 degenerate base in the sequence or, in other words, won't consider the areas

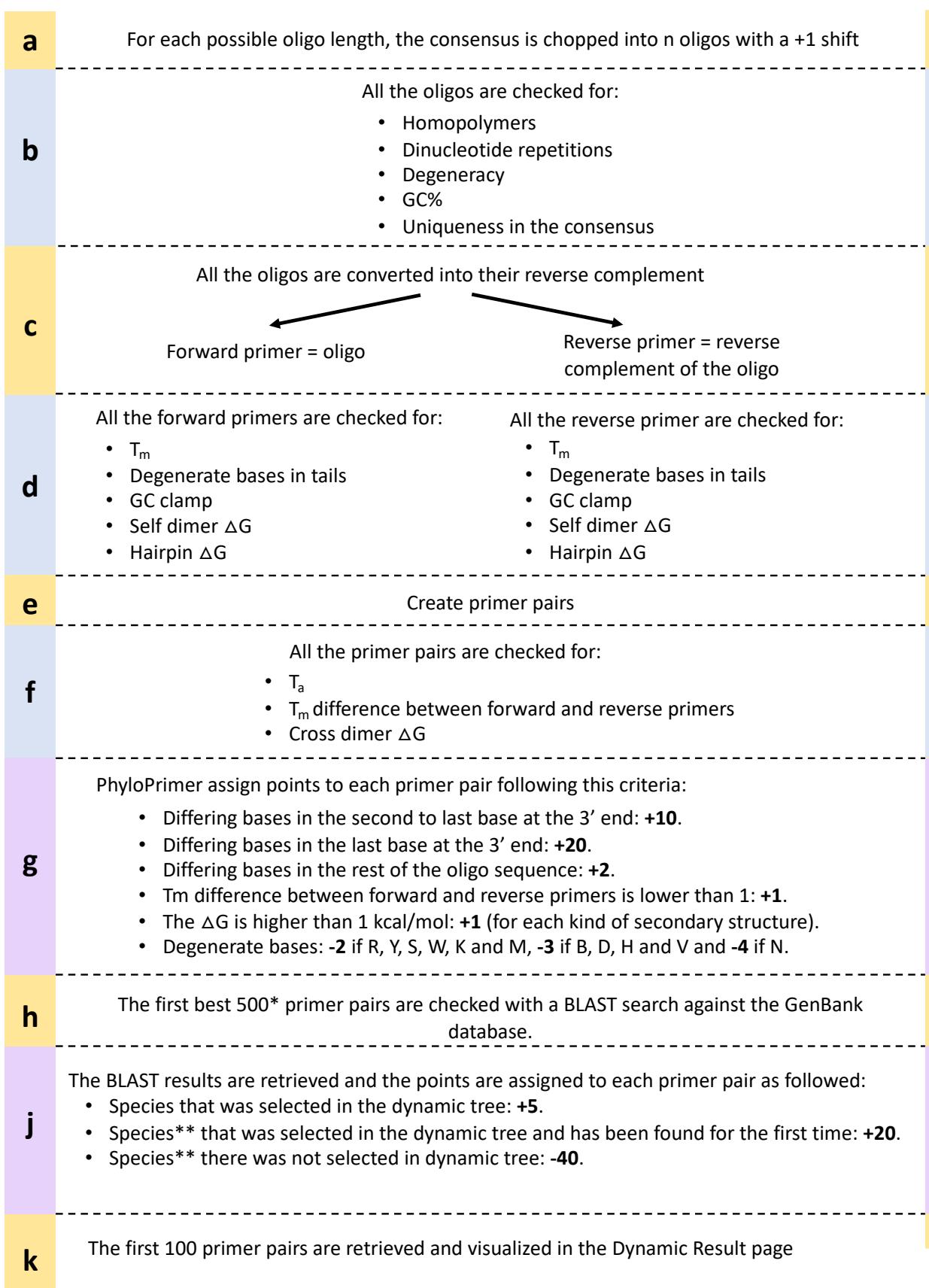


Figure 8. Primer design workflow. The yellow rectangles define the design processes, the blues rectangles define the check processes and the purple defined the check processes. *250 if no negative consensus was present, no differing bases between the two consensus sequences were present or no differences were taken in consideration in the scoring system. **depending on the visualization criteria that were selected, +20 and -40 points are assigned if the different oligos were BLASTed against GenBank entries belonging to genera, families, orders, classes, phyla and domains that were attributed to the sequences selected from the phylogenetic tree.

of the consensus that have an incidence of degenerate bases higher than 1 base every 18-22 bases (the oligo length by default). If the maximum degenerate base value was set to 2, PhyloPrimer will discard the oligos and therefore the consensus area that had more than 2 bases every 18-22 bases, and so on.

PhyloPrimer will start the oligo design only once the positive consensus has been obtained (Figure 8). For each possible oligo length (between 18 and 22 bases by default), the software extracts from the consensus sequence all the possible subsequences of that length. This first step creates the starting pool of oligos that the following steps will check and discard if not respecting all the design parameters set in the Oligo Design pages. The first check step discards by default the oligos that are not unique in the consensus sequence, that have homopolymer repetition longer than 3 bases, dinucleotide repetition longer than 6 bases, a GC content lower than 40% or higher than 60%, and will check and discard the oligos that do not have between 2 and 4 Gs/Cs in the last 5 bases of 3' oligo end (GC clamp). PhyloPrimer will also check if the oligos have a higher number of degenerate bases than the limit and that only the correct degenerate bases are present (all except from N by default). The default number of degenerate bases is set by PhyloPrimer in relation to how many degenerate bases were found inside the consensus sequence but can be changed by the user.

PhyloPrimer then calculates the reverse complement of all the oligos and considers the original oligos as putative forward primers and the oligo reverse complements as putative reverse primers. All the forward and reverse primers are progressively checked to have a valid melting temperature (between 54 °C and 64 °C by default) and, in case the presence of degenerate bases is allowed, not to have degenerate bases in the last 5 bases of the 5' oligo end and last 2 bases of the 3' end oligo tails (by default). The software also checks for the presence of self dimer and hairpin secondary structures and discards any oligos with a secondary structure associated to a ΔG value lower than -5 and -3 kcal mol⁻¹, respectively (by default). If the primers had values that reflected the set criteria, PhyloPrimer uses them for finding suitable primer pairs. The primer pairs are first selected considering the distance between their 5' ends on the consensus (between 200 and 600 bases by default). All the suitable primer pairs are then checked for the presence of cross dimer formations and discarded if the ΔG values are lower than -5 kcal mol⁻¹. Furthermore the primer pairs are also discarded if the melting temperature difference between forward and reverse primers is higher than 5 °C or the annealing temperature does not range between 50 °C and 60 °C (by default).

At this point, all the remaining primer pairs have all the requirements that were set by the user through the Oligo Design pages. All the following steps aim to retrieve the best primer pairs that will be visualized in the Dynamic Result page. This is achieved assigning points to each primer pair as follows: 1 point is assigned to the primer pair if the melting temperature of the forward and reverse primers differed for less than 1 °C, for each secondary structure 1 point is assigned if the ΔG value is higher than -1 kcal mol⁻¹. Moreover, 20 points if a differing base is present in the last base of the 3' end and 10 points if it is present in the second to last position of the oligo 3' end, 2 points are assigned for each other differing bases between the positive and the negative consensus.

PhyloPrimer is then going to select the first 500 primer pairs that scored the highest points according to the scoring system. If working with primer pairs or primer-probe assays, PhyloPrimer will select only 250 primer pairs if no negative consensus is present, no differences are present between positive and negative consensus or the user does not want PhyloPrimer to assign points to oligos that fall inside differing position between the two consensus sequences. In all the latter cases, in fact, the oligos would be selected on all the consensus sequence and not localized on a specific consensus area and the oligos, the oligos would be different between each other and PhyloPrimer would take too long to check them. The oligos belonging to those first 500 primer pairs will be BLASTed against the GenBank database. PhyloPrimer is then going to check the BLAST results and it is going to consider only the BLAST matches that were matched by both the forward and the reverse primers of a primer pairs, if that sequence belongs to one of the species that were selected from the dynamic tree, PhyloPrimer is going to assign 10 points to the primer pair, if the species was not among the selected species it will deduct 40 points and every time there is a new correct species is going to add 20 points to the total. By default, PhyloPrimer will not assign more points to primers that belong to the same genus (or higher ranks) of the selected tree entries. But if these visualization parameters are checked, PhyloPrimer will assign 20 points to the entries that belonging to the same taxonomy and deduct 40 to those that do not. This is for facilitating the design of oligos that are specific for a specific genus (or higher taxonomic group) rather than only specific to certain species. In case an additional file was uploaded by the user for an additional BLAST check, PhyloPrimer will also blast all the oligos against that database but the outcome will not be object of the scoring system.

The described scoring criteria are all active by default but any of those can be deselected by the user on the Oligo Design page. PhyloPrimer then selects the first 100 primer pairs and these primer pairs will be the ones showed in the last Result Page. When degenerate bases are present inside the oligo sequences, the melting temperature and the GC content are calculated as the mean of these values in each of the possible oligo.

Above we described the process for the design of primer pairs but it is very similar to the workflow to design primer pair/probe assays and single oligos. In case a single oligo was to be designed, the workflow would be the same as for the primer design but no oligo pairs would be formed and consequently none of the checks for it. The following modification to the workflow would be used in case a primer pair/probe assays was to be designed:

- In case of the designing of primer pair/probe assays PhyloPrimer would have also designed the probe. The

latter is designed with the same process as for the forward primer in case it is an anti-sense probe, as for the reverse primers in case it is a sense probe or as for both the primers if the binding DNA template strand is not specified.

- Probe sequences with a guanine in the last base of the 5' end are discarded.
- During the process of creating the primer pairs, PhyloPrimer would have also looked for a probe that falls inside a consensus region between the two primers.
- The primer pair/probe assays are checked for the melting temperature difference between the probe and the primers. The assay is discarded if the difference between the melting temperature of the probe and the one of the primers (the average between forward and reverse primers) is lower than 10 °C (by default).

Finally, the check mode the checks shown in Figure 8d-f would be performed (varying in relation to which kind of oligo was uploaded by the user).

5 Consensus calculation

The consensus is a nucleotide sequence calculated from alignment sequences and reports the most frequent base for each alignment position (Figure 9a). Alignment sequences have all the same lengths. The software goes through it position by position (or column by column) and for each position looks at the base composition for all the alignments. For each position, 1 point is assigned to each A, C, G, T or gap (- or .) that is found. When there are degenerate bases in the alignment sequences, fractionated points are assigned to the correspondent bases of the degenerate base (Table 1). For instance, if the degenerate base N is present, 1/4 point is assigned to A, 1/4 to C, 1/4 to T and 1/4 to G; in case of a B , 1/3 is assigned to G, 1/3 to C and 1/3 to T; if S is present 1/2 point is assigned to C and 1/2 to G. Once the software knows the frequency of each of the four base (and the gap) at all the positions, it determines which base to assign to each position following the criteria illustrated in the table in Figure 7b. For example, if for a certain position, 50% or more of the alignment sequences had a gap, a gap will be reported in the consensus, if one of the bases represents the 20% or more and each of the other less than 20%, that position will be assigned to that letter. If there are more than two bases present with 20% or more abundance, that position will be assigned to the corresponding degenerate base.

6 Thermodynamic calculation

Oligo melting temperatures (T_m) and Gibbs free energies (ΔG) of all the possible oligo secondary structures are calculated with the nearest-neighbor (NN) model for nucleic acids. This model predicts the thermodynamic behavior of a DNA molecule using the thermodynamic parameters of each nucleotide pair composing the molecule itself. The ΔG is used to estimate how stable a particular DNA structure is at a certain temperature. The lower the ΔG the more probable is a certain secondary structure (i.e. cross-dimers, self-dimers and hairpins) to occur. Both the T_m and the ΔG calculation rely on the use of the thermodynamics parameters enthalpy (ΔH) and entropy (ΔS). These parameters were derived from calorimetry and spectroscopic experiments of DNA duplexes for nucleotide base pair motives (SantaLucia and Hicks, 2004), internal mismatches (Allawi and SantaLucia, 1997; Allawi and SantaLucia, 1998a; Allawi and SantaLucia, 1998b; Allawi and SantaLucia, 1998c; Peyret et al., 1999), dangling ends (Bommarito et al., 2000) and hairpin terminal mismatches (unpublished data). The latter were retrieved from the UNAFold database (Markham and Zuker, 2008). The ΔH and ΔS are considered temperature independent when working with nucleic acids and are reported for 1M Na⁺ conditions. As the PCR conditions can span a wide range of different conditions, we need to apply salt correction formulas (Owczarzy et al., 2004; Owczarzy et al., 2008). Furthermore, other correction formulas need to be applied when working with DNA duplexes that present internal loops and hairpin formations (SantaLucia and Hicks, 2004).

6.1 Melting temperature

The melting temperature (T_m) of a DNA molecule is the temperature in which half of the DNA is paired with its complement and half is single stranded. The correct calculation of this parameter is essential to the correct calculation of the PCR annealing temperature, and it is pivotal for the qPCR probe when wanting to differentiate amplicon expression levels. T_m is calculated with the formula reported from SantaLucia and Hicks (2004). The following formula calculates the T_m at 1 M Na⁺ condition.

a) alignments:

| | | |
|--|--|---|
| <pre> 1 2 3 4 5 6 7 8 9 TAGGNNSGC- TTGCCACC- TTGCCACC- TTGCCAGCA TTGCCAAC </pre> | | <pre> 1. T:5 2. A:1 T:4 3. G:5 4. G:2 C:3 5. A:0.25 T:0.25 G:0.25 C:4.25 6. A:4 G:0.5 C:0.5 7. A:1 G:2 C:2 8. C:5 9. A:2 gap:3 </pre> |
|--|--|---|

consensus:

| | |
|--|--|
| <pre> 1 2 3 4 5 6 7 8 9 TWGCCASC- </pre> | |
|--|--|

| |
|---|
| <pre> 1. T:1 → T 2. A:0.2 T:0.8 → W 3. G:1 → G 4. G:0.4 C:0.6 → C 5. A:0.05 T:0.5 G:0.05 C:0.85 → C 6. A:0.8 G:0.1 C:0.1 → A 7. A:0.2 G:0.4 C:0.4 → S 8. C:1 → C 9. A: 0.4 gap:0.6 → - </pre> |
|---|

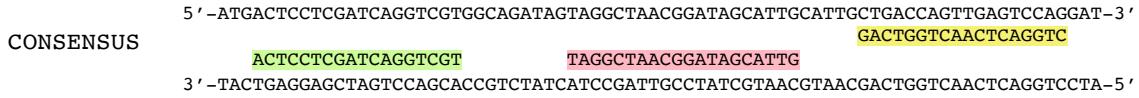
b)

| | | Bases at a specific position in the alignments | | | | |
|---------------------------------|-----|--|------|------|------|------|
| | | A | T | C | G | gap |
| Assigned bases in the consensus | A | ≥ 20 | < 20 | < 20 | < 20 | |
| | T | < 20 | ≥ 20 | < 20 | < 20 | |
| | C | < 20 | < 20 | ≥ 20 | < 20 | |
| | G | < 20 | < 20 | < 20 | ≥ 20 | |
| | R | ≥ 20 | < 20 | < 20 | ≥ 20 | |
| | Y | < 20 | ≥ 20 | ≥ 20 | < 20 | |
| | S | < 20 | < 20 | ≥ 20 | ≥ 20 | |
| | W | ≥ 20 | ≥ 20 | < 20 | < 20 | |
| | K | < 20 | ≥ 20 | < 20 | ≥ 20 | |
| | M | ≥ 20 | < 20 | ≥ 20 | < 20 | |
| | B | < 20 | ≥ 20 | ≥ 20 | ≥ 20 | |
| | D | ≥ 20 | ≥ 20 | < 20 | ≥ 20 | |
| | H | ≥ 20 | ≥ 20 | ≥ 20 | < 20 | |
| | V | ≥ 20 | < 20 | ≥ 20 | ≥ 20 | |
| | N | ≥ 20 | ≥ 20 | ≥ 20 | ≥ 20 | ≥ 50 |
| | gap | | | | | |

Figure 9. The consensus calculation workflow. a) represents how PhyloPrimer processes alignment sequences into a consensus sequence whereas b) shows which bases are assigned to each consensus position in relation to which base frequencies were in each consensus position.

$$T_m(1 \text{ M Na}^+) = 1000 \times \frac{\Delta H^o}{\Delta S^o + R \ln(\frac{C_x}{x})} \quad (1)$$

where:



a) FORWARD PRIMER 5' -**ACTCCTCGATCAGGTCTG** -3' forward primer
 3' -TGAGGAGCTAGTCCAGCAC-5' consensus

$$\begin{aligned}
 \Delta H &= \Delta H(\text{initiation})^{*1} + \Delta H(\text{terminal_AT})^{*2} + \Delta H(\text{AC/TG})^{*3} + \Delta H(\text{CT/GA})^{*3} + \Delta H(\text{TC/AG})^{*3} + \Delta H(\text{CC/GG})^{*3} + \Delta H(\text{CT/GA})^{*3} + \\
 &\Delta H(\text{TC/AG})^{*3} + \Delta H(\text{CG/GC})^{*3} + \Delta H(\text{GA/CT})^{*3} + \Delta H(\text{AT/TA})^{*3} + \Delta H(\text{TC/AG})^{*3} + \Delta H(\text{CA/GT})^{*3} + \Delta H(\text{AG/TC})^{*3} + \Delta H(\text{GG/CC})^{*3} + \\
 &\Delta H(\text{GT/CA})^{*3} + \Delta H(\text{TC/AG})^{*3} + \Delta H(\text{CG/GC})^{*3} + \Delta H(\text{GT/CA})^{*3} + \Delta H(\text{TX/AC})^{*3} = +0.2 +2.2 -8.4 -7.8 -8.2 -8 -7.8 -8.2 - \\
 &10.6 -8.2 -7.2 -8.2 -8.5 -7.8 -8 -8.4 -10.6 -8.4 +0.6 = -139.5 \text{ kcal mol}^{-1} \\
 \Delta S &= \Delta S(\text{initiation})^{*1} + \Delta S(\text{terminal_AT})^{*2} + \Delta S(\text{AC/TG})^{*3} + \Delta S(\text{CT/GA})^{*3} + \Delta S(\text{TC/AG})^{*3} + \Delta S(\text{CC/GG})^{*3} + \Delta S(\text{CT/GA})^{*3} + \\
 &\Delta S(\text{TC/AG})^{*3} + \Delta S(\text{CG/GC})^{*3} + \Delta S(\text{GA/CT})^{*3} + \Delta S(\text{AT/TA})^{*3} + \Delta S(\text{TC/AG})^{*3} + \Delta S(\text{CA/GT})^{*3} + \Delta S(\text{AG/TC})^{*3} + \Delta S(\text{GG/CC})^{*3} + \\
 &\Delta S(\text{GT/CA})^{*3} + \Delta S(\text{TC/AG})^{*3} + \Delta S(\text{CG/GC})^{*3} + \Delta S(\text{GT/CA})^{*3} + \Delta S(\text{TX/AC})^{*3} = -5.7 +6.9 -22.4 -21 -22.2 -19.9 -21 -22.2 \\
 &-27.2 -22.2 -20.4 -22.2 -22.7 -21 -19.9 -22.4 -22.2 -27.2 -22.4 +3.3 = -374 \text{ cal K}^{-1} \text{ mol}^{-1} \\
 T_m(1 \text{ M Na}^+) &= 1000 \times (-139.5 / (-374 + 1.9872 \times \ln(0.000002/4))) = 346.3 \text{ K (73.1 }^\circ\text{C)}
 \end{aligned}$$

b) REVERSE PRIMER 5' -**CTGGACTCAACTGGTCAG** -3' reverse primer
 3' -GACCTGAGTTGACCGAGTCG-5' consensus

$$\begin{aligned}
 \Delta H &= \Delta H(\text{initiation})^{*1} + \Delta H(\text{CT/GA})^{*3} + \Delta H(\text{TG/AC})^{*3} + \Delta H(\text{GG/CC})^{*3} + \Delta H(\text{GA/CT})^{*3} + \Delta H(\text{AC/TG})^{*3} + \Delta H(\text{CT/GA})^{*3} + \\
 &\Delta H(\text{TC/AG})^{*3} + \Delta H(\text{CA/GT})^{*3} + \Delta H(\text{AA/TT})^{*3} + \Delta H(\text{AC/TG})^{*3} + \Delta H(\text{CT/GA})^{*3} + \Delta H(\text{TG/AC})^{*3} + \Delta H(\text{GG/CC})^{*3} + \Delta H(\text{GT/CA})^{*3} + \\
 &\Delta H(\text{TC/AG})^{*3} + \Delta H(\text{CA/GT})^{*3} + \Delta H(\text{AG/TC})^{*3} + \Delta H(\text{GX/GC})^{*3} = +0.2 -7.8 -8.5 -8 -8.2 -8.4 -7.8 -8.2 -8.5 -7.6 -8.4 - \\
 &7.8 -8.5 -8 -8.4 -8.2 -8.5 -7.8 -5.1 = -143.5 \text{ kcal mol}^{-1} \\
 \Delta S &= \Delta S(\text{initiation})^{*1} + \Delta S(\text{CT/GA})^{*3} + \Delta S(\text{TG/AC})^{*3} + \Delta S(\text{GG/CC})^{*3} + \Delta S(\text{GA/CT})^{*3} + \Delta S(\text{AC/TG})^{*3} + \Delta S(\text{CT/GA})^{*3} + \\
 &\Delta S(\text{TC/AG})^{*3} + \Delta S(\text{CA/GT})^{*3} + \Delta S(\text{AA/TT})^{*3} + \Delta S(\text{AC/TG})^{*3} + \Delta S(\text{CT/GA})^{*3} + \Delta S(\text{TG/AC})^{*3} + \Delta S(\text{GG/CC})^{*3} + \Delta S(\text{GT/CA})^{*3} + \\
 &\Delta S(\text{TC/AG})^{*3} + \Delta S(\text{CA/GT})^{*3} + \Delta S(\text{AG/TC})^{*3} + \Delta S(\text{GX/GC})^{*3} = -5.7 -21 -22.7 -19.9 -22.2 -22.4 -21 -22.2 -22.7 -21.3 - \\
 &22.4 -21 -22.7 -19.9 -22.4 -22.2 -22.7 -21 -14 = -389.4 \text{ cal K}^{-1} \text{ mol}^{-1} \\
 T_m(1 \text{ M Na}^+) &= 1000 \times (-143.5 / (-389.4 + 1.9872 \times \ln(0.000002/4))) = 343.1 \text{ K (69.95 }^\circ\text{C)}
 \end{aligned}$$

c) PROBE 5' - **TAGGCTAACGGATAGCATTG** -3' anti-sense probe
 3' -CATCCGATTGCCATATCGTAACG-5' consensus

$$\begin{aligned}
 \Delta H &= \Delta H(\text{initiation})^{*1} + \Delta H(\text{XT_CA})^{*3} + \Delta H(\text{TA/AT})^{*3} + \Delta H(\text{AG/TC})^{*3} + \Delta H(\text{GG/CC})^{*3} + \Delta H(\text{GC/CN})^{*3} + \Delta H(\text{CT/GA})^{*3} + \\
 &\Delta H(\text{TA/AT})^{*3} + \Delta H(\text{AA/TT})^{*3} + \Delta H(\text{AC/TG})^{*3} + \Delta H(\text{CG/GC})^{*3} + \Delta H(\text{GG/CC})^{*3} + \Delta H(\text{GA/CT})^{*3} + \Delta H(\text{AT/TA})^{*3} + \Delta H(\text{TA/AT})^{*3} + \\
 &\Delta H(\text{AG/TC})^{*3} + \Delta H(\text{GC/CN})^{*3} + \Delta H(\text{CA/GT})^{*3} + \Delta H(\text{AT/TA})^{*3} + \Delta H(\text{TT/AA})^{*3} + \Delta H(\text{TG/AC})^{*3} + \Delta H(\text{GX/CT})^{*3} = +0.2 +4.7 -7.2 \\
 &-7.8 -8 -9.8 -7.8 -7.2 -7.6 -8.4 -10.6 -8 -8.2 -7.2 -7.2 -7.8 -9.8 -8.5 -7.2 -7.6 -8.5 -5.1 = -154.6 \text{ kcal mol}^{-1} \\
 \Delta S &= \Delta S(\text{initiation})^{*1} + \Delta S(\text{XT/CA})^{*3} + \Delta S(\text{TA/AT})^{*3} + \Delta S(\text{AG/TC})^{*3} + \Delta S(\text{GG/CC})^{*3} + \Delta S(\text{GC/CN})^{*3} + \Delta S(\text{CT/GA})^{*3} + \\
 &\Delta S(\text{TA/AT})^{*3} + \Delta S(\text{AA/TT})^{*3} + \Delta S(\text{AC/TG})^{*3} + \Delta S(\text{CG/GC})^{*3} + \Delta S(\text{GG/CC})^{*3} + \Delta S(\text{GA/CT})^{*3} + \Delta S(\text{AT/TA})^{*3} + \Delta S(\text{TA/AT})^{*3} + \\
 &\Delta S(\text{AG/TC})^{*3} + \Delta S(\text{GC/CN})^{*3} + \Delta S(\text{CA/GT})^{*3} + \Delta S(\text{AT/TA})^{*3} + \Delta S(\text{TT/AA})^{*3} + \Delta S(\text{TG/AC})^{*3} + \Delta S(\text{GX/CT})^{*3} = -5.7 +14.2 - \\
 &21.3 -21 -19.9 -24.4 -21 -21.3 -21.3 -22.4 -27.2 -19.9 -22.2 -20.4 -21.3 -21 -24.4 -22.7 -20.4 -21.3 -22.7 -14 \\
 &= -421.6 \text{ cal K}^{-1} \text{ mol}^{-1} \\
 T_m(1 \text{ M Na}^+) &= 1000 \times (-154.6 / (-421.6 + 1.9872 \times \ln(0.000002/4))) = 343.2 \text{ K (70.05 }^\circ\text{C)}
 \end{aligned}$$

Figure 10. A schematic view of how the melting temperature is calculated for a) forward primers, b) reverse primers and c) probes. ΔH^{*1} and ΔS^{*1} refer to duplex initiation values; ΔH^{*2} and ΔS^{*2} refer to A and T terminals; ΔH^{*3} and ΔS^{*3} refer to NN base pairs.

$T_m(1 \text{ M Na}^+)$ = melting temperature at 1 M Na⁺ conditions (K).

ΔH° = enthalpy (kcal mol⁻¹).

R = gas constant, 1.9872 (cal K⁻¹ mol⁻¹).

ΔS° = entropy (cal K⁻¹ mol⁻¹).

C_T = total molar strand concentration (mol L⁻¹). It is approximated to the concentration of the oligos (forward, reverse and probe concentration). It does not consider the template concentration because it is supposed to be significantly lower than the oligo concentration.

x = 4 for nonself-complementary duplexes (e.g. primers) and 1 for self-complementary duplexes.

Figure 10 shows how the equation 1 is applied to the calculation of the T_m for forward primers (a), reverse primers (b) and probes (c). In order to apply the equation 1, we need first to calculate ΔH and ΔS which are the sum of all the ΔH and ΔS that influence the annealing temperature of the oligos to the DNA template. For each value, the following components are separately summed up:

- ΔH and ΔS initiation values which consider the fact that duplex initiation is not likely to happen, the ΔG is in fact positive (marked as ΔH^{*1} and ΔS^{*1} in the Figure 10; Table A1a).
- ΔH and ΔS terminal values in case the oligo duplex ends with an adenine or a thymine (marked as ΔH^{*2} and ΔS^{*2} in the Figure 10; Table A1a).
- ΔH and ΔS values for all the nearest-neighbor base pairs (marked as ΔH^{*3} and ΔS^{*3} in the Figure 10; Table A1a-c).

Also, because of how the PCR works, the forward and the reverse primers are considered to have a blunt-end (e.g. $\frac{A}{T} G$) at the 5'-end terminal and a dangling end (e.g. $\frac{A}{T} C$; Table A1c) at the 3' terminal (Figure 10a-b). The probes are considered to terminate with both dangling ends (Figure 10c). This is an approximation as i) for the first PCR steps the primer's ends terminate both in dangling ends and ii) the dangling base depends on the DNA template the oligo anneals to. PhyloPrimer can predict the dangling ends from the consensus sequence as it was used for the oligo design but the DNA template may vary during the PCR. In the Oligo Check mode PhyloPrimer considers all oligos to end with two blunt-ends as it does not know the DNA template. When degenerate bases are present in the DNA duplexes, the T_m is calculated for all the oligo variants and only the average T_m is reported.

The annealing temperature, T_a , is calculated as the lowest melting temperature (if more than one oligo is present) minus 5. This is an indicative calculation as the annealing temperature can considerably vary in relation to the polymerase that is used during the PCR.

6.2 ΔG calculation

The ΔG , or Gibbs free energy, indicates if a reaction can occur spontaneously (ΔG higher than 0, exergonic reaction) or not (ΔG lower than 0, endergonic reaction). In our case, ΔG represents the quantity of energy needed to fully break a secondary structure. The lower it is (more negative), the more stable and likely to occur the secondary structure will be and the more energy will be required to break it. ΔG is defined as equal to the enthalpy minus the product of the temperature times the entropy:

$$\Delta G_T^o = \Delta H^o - T\Delta S^o \quad (2)$$

where:

ΔG^o = Gibbs free energy (kcal mol⁻¹)

T = temperature (K). Ideally this values should be set to the one that is used during the extension phase of the amplification. The two standard values used for the ΔG calculation are 298.15 K (25 °C) and 308.15 K (37 °C). However, PhyloPrimer sets 72 °C as default value as it corresponds to a common annealing temperature in the PCR.

From this formula we can observe that ΔG is temperature dependent and that ΔH and ΔS are temperature independent, which is an approximation used when working with nucleic acids. The three different secondary structures PhyloPrimer considers are self-dimers, which are dimers formed within the oligo itself, cross-dimers, which are dimers formed between different oligos, and hairpin loops which are hairpin-like secondary structures formed within the oligo itself. For each secondary structure, the ΔG is calculated on the longest stretch of consecutive bases containing only single mismatches. If an oligo presents more than one stretch that reflects this property, the ΔG will be calculated for each stretch and the final ΔG will be attributed to the lowest ΔG values (Figure 11b-f). As seen for the T_m calculation, before applying the equation 2, ΔH and ΔS must be calculated first. These two values are obtained by the sum of different aspects depending which secondary structure is taken into analyses. Enthalpies and entropies for self-dimers (Figure 11a-b) and cross-dimers (Figure 11c) are calculated as follows:

- ΔH and ΔS initiation values which consider the fact that duplex initiation is not likely to happen, the ΔG is positive (marked as ΔH^{*1} and ΔS^{*1} in the Figure 11; Table A1a).
- ΔH and ΔS terminal values in case the oligo duplex ends with an adenine or a thymine (marked as ΔH^{*2} and ΔS^{*2} in the Figure 11; Table A1a).
- ΔH and ΔS values for all the nearest-neighbor base pairs (marked as ΔH^{*3} and ΔS^{*3} in the Figure 10; Table A1a-c).
- ΔH and ΔS values that accounts for symmetry (marked as ΔH^{*4} and ΔS^{*4} in the Figure 11; Table A1a). They are applied only to self dimers.

When oligos with degenerate bases (e.g. ATASTG) were analyzed by PhyloPrimer the self dimer formations are calculated following the self dimer rules outlined above when considering the secondary structures formed within the same oligo variety (e.g. ATACTG vs ATACTG) and following the cross dimer rules when considering the secondary structures formed between different oligo varieties (e.g. ATACTG vs ATAGTG). The latter is reported as self dimer formations by PhyloPrimer but is marked with an * in the **Result Page** (Section 2.4.2).

In the case of the ΔG calculation for hairpins loops there are different aspects to take in consideration which vary depending on the length of the hairpin loop. Hairpin loops of less than 3 bases are not considered thermodynamically stable and therefore are not taken in consideration. Furthermore, certain hairpin loops that are 3 or 4 bases are

particularly stable and thermodynamics corrections are applied to them. In case the hairpin loop is 3-base long (Figure 11d), PhyloPrimer takes in consideration the following ΔH and ΔS values:

- ΔH and ΔS initiation values which consider the fact that duplex initiation is not likely to happen, the ΔG is positive (marked as ΔH^{*1} and ΔS^{*1} in the Figure 11; Table A1a).
- ΔH and ΔS terminal values in case the oligo duplex ends with an adenine or a thymine at both the duplex extremities (marked as ΔH^{*2} and ΔS^{*2} in the Figure 11; Table A1a).
- ΔH and ΔS values for all the nearest-neighbor base pairs (marked as ΔH^{*3} and ΔS^{*3} in the Figure 11; Table A1a-c).
- ΔH and ΔS values for 3-base long loops (marked as ΔH^{*5} and ΔS^{*5} in the Figure 11; Table A2a).
- ΔH and ΔS bonus values for particular 3-base loop sequences (marked as ΔH^{*6} and ΔS^{*6} in the Figure 11; Table A2b).

In case the hairpin loop is 4-base long (Figure 11e), PhyloPrimer takes in consideration the following ΔH and ΔS values:

- ΔH and ΔS initiation values which consider the fact that duplex initiation is not likely to happen, the ΔG is positive (marked as ΔH^{*1} and ΔS^{*1} in the Figure 11; Table A1a).
- ΔH and ΔS terminal values in case the oligo duplex ends with an adenine or a thymine not at the loop extremities (marked as ΔH^{*2} and ΔS^{*2} in the Figure 11; Table A1a).
- ΔH and ΔS values for all the nearest-neighbor base pairs (marked as ΔH^{*3} and ΔS^{*3} in the Figure 11; Table A1a-c).
- ΔH and ΔS values for 4-base long loops (marked as ΔH^{*5} and ΔS^{*5} in the Figure 11; Table A2a).
- ΔH and ΔS bonus values for particular 4-base loop sequences (marked as ΔH^{*6} and ΔS^{*6} in the Figure 11; Table A2b).
- ΔH and ΔS values on the terminal pair that closes to the loop (marked as ΔH^{*7} and ΔS^{*7} in the Figure 11; Table A3).

In case the hairpin loop is longer than 4 bases (Figure 11f), PhyloPrimer takes in consideration the following ΔH and ΔS values:

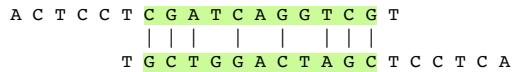
- ΔH and ΔS initiation values which consider the fact that duplex initiation is not likely to happen, the ΔG is positive (marked as ΔH^{*1} and ΔS^{*1} in the Figure 11; Table A1a).
- ΔH and ΔS terminal values in case the oligo duplex ends with an adenine or a thymine not at the loop extremities (marked as ΔH^{*2} and ΔS^{*2} in the Figure 11; Table A1a).
- ΔH and ΔS values for all the nearest-neighbor base pairs (marked as ΔH^{*3} and ΔS^{*3} in the Figure 11; Table A1a-c).
- ΔH and ΔS values for n-base long loops (marked as ΔH^{*5} and ΔS^{*5} in the Figure 11; Table A2a).
- ΔH and ΔS values on the terminal pair that closes to the loop (marked as ΔH^{*7} and ΔS^{*7} in the Figure 11; Table A3).

The hairpin-like duplex will have 2 ends: the one terminating into the loop and the other one (on the other extremity). As we saw above, the loop termination is treated differently depending on the loop size, the other end behaves as we saw for the self- and cross-dimers where it can be a blunt-end or a dangling end. When an oligo has degenerate bases in its sequences, the ΔG is calculated for all the possible oligos and all the ΔG will be reported (when lower than 0 kcal mol⁻¹).

6.3 Formula corrections

All the thermodynamic parameters used in Figure 11 and 12 and reported in Table 2, 3, 4 and 5 are obtained for 1 M Na⁺ condition. Depending on the used polymerase and the PCR protocol, Mg²⁺ and monovalent ions can vary considerably and rarely the 1 M Na⁺ condition is respected. Mg²⁺ can be added to the PCR reaction as MgCl₂ or MgSO₄, it is a DNA polymerase co-factor and it helps the stabilization of the primer-template DNA duplex influencing the negative charges of the DNA backbones. Monovalent ions such as Na⁺ and K⁺ (KCl) are also used for stabilizing the DNA duplex, whereas Tris-HCl is used for stabilizing the pH. An high amount of these ions leads to a low oligo specificity, increasing the melting temperature. As it has been shown as the presence of monovalent and divalent ions influence the T_m and the ΔG calculation, in the following sections we introduce the

a) SELF-DIMERS – case 1

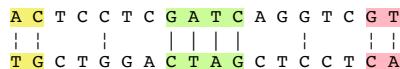


$$\Delta H = \Delta H(\text{initiation})^{*1} + \Delta H(CG_GC)^{*3} + \Delta H(GA_CT)^{*3} + \Delta H(AT_TG)^{*3} + \Delta H(TC_GG)^{*3} + \Delta H(CA_GA)^{*3} + \Delta H(AG_AC)^{*3} + \Delta H(GG_CT)^{*3} + \Delta H(GT_TA)^{*3} + \Delta H(TC_AG)^{*3} + \Delta H(CG_GC)^{*3} + \Delta H(\text{symmetry})^{*4} = +0.2 - 10.6 - 8.2 - 2.5 + 3.3 - 0.9 - 0.9 + 3.3 - 2.5 - 8.2 - 10.6 + 0 = -37.6 \text{ kcal mol}^{-1}$$

$$\Delta S = \Delta S(\text{initiation})^{*1} + \Delta S(CG_GC)^{*3} + \Delta S(GA_CT)^{*3} + \Delta S(AT_TG)^{*3} + \Delta S(TC_GG)^{*3} + \Delta S(CA_GA)^{*3} + \Delta S(AG_AC)^{*3} + \Delta S(GG_CT)^{*3} + \Delta S(GT_TA)^{*3} + \Delta S(TC_AG)^{*3} + \Delta S(CG_GC)^{*3} + \Delta S(\text{symmetry})^{*4} = -5.7 - 27.2 - 22.2 - 8.3 + 10.4 - 4.2 - 4.2 + 10.4 - 8.3 - 22.2 - 27.2 - 1.4 = 110.1 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -37.6 - (298.15 \times -110.1/1000) = -4.77 \text{ kcal mol}^{-1}$$

b) SELF-DIMERS – case 2



$$\Delta H = \Delta H = \Delta H(\text{initiation})^{*1} + \Delta H(AC_TG)^{*3} + \Delta H(\text{terminal_AT})^{*3} + \Delta H(\text{symmetry})^{*4} = +0.2 - 8.4 + 2.2 + 0 = -6 \text{ kcal mol}^{-1}$$

$$\Delta S = \Delta S = \Delta S(\text{initiation})^{*1} + \Delta S(AC_TG)^{*3} + \Delta S(\text{terminal_AT})^{*3} + \Delta S(\text{symmetry})^{*4} = -5.7 - 22.4 + 6.9 - 1.4 = -22.6 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = \Delta G = -6 - (298.15 \times -22.6/1000) = +0.73 \text{ kcal mol}^{-1}$$

$$\Delta H = \Delta H(\text{initiation})^{*1} + \Delta H(GA_CT)^{*3} + \Delta H(AT_TA)^{*3} + \Delta H(TC_AG)^{*3} + \Delta H(\text{symmetry})^{*4} = +0.2 - 8.2 - 7.2 - 8.2 + 0 = -23.4 \text{ kcal mol}^{-1}$$

$$\Delta S = \Delta S(\text{initiation})^{*1} + \Delta S(GA_CT)^{*3} + \Delta S(AT_TA)^{*3} + \Delta S(TC_AG)^{*3} + \Delta S(\text{symmetry})^{*4} = -5.7 - 22.2 - 20.4 - 22.2 - 1.4 = -71.9 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -23.4 - (298.15 \times -71.9/1000) = -1.9 \text{ kcal mol}^{-1}$$

$$\Delta G = \Delta G = -1.9 \text{ kcal mol}^{-1}$$

c) CROSS-DIMERS – case 1

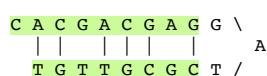


$$\Delta H = \Delta H(\text{initiation})^{*1} + \Delta H(\text{terminal_AT})^{*2} + \Delta H(TC_AG)^{*3} + \Delta H(CG_GC)^{*3} + \Delta H(GA_CT)^{*3} + \Delta H(AC_TG)^{*3} + \Delta H(CG_GC)^{*3} = +0.2 + 2.2 - 8.2 - 10.6 - 8.2 - 8.4 - 10.6 = -43.6 \text{ kcal mol}^{-1}$$

$$\Delta S = \Delta S(\text{initiation})^{*1} + \Delta S(\text{terminal_AT})^{*2} + \Delta S(TC_AG)^{*3} + \Delta S(CG_GC)^{*3} + \Delta S(GA_CT)^{*3} + \Delta S(AC_TG)^{*3} + \Delta S(CG_GC)^{*3} = -5.7 + 6.9 - 22.2 - 27.2 - 22.2 - 22.4 - 27.2 = -120 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -43.6 - (298.15 \times -120/1000) = -7.8 \text{ kcal mol}^{-1}$$

d) HAIRPINS – case 1

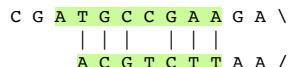


$$\Delta H = \Delta H(CA_XT)^{*3} + \Delta H(AC_TG)^{*3} + \Delta H(CG_GT)^{*3} + \Delta H(GA_TT)^{*3} + \Delta H(AC_TG)^{*3} + \Delta H(CG_GC)^{*3} + \Delta H(\text{loop_3})^{*5} + \Delta H(GGATC)^{*3} = +0.6 - 8.4 - 4.1 - 1.3 - 8.4 - 10.6 + 0 + 0 = -32.2 \text{ kcal mol}^{-1}$$

$$\Delta S = \Delta S(CA_XT)^{*3} + \Delta S(AC_TG)^{*3} + \Delta S(CG_GT)^{*3} + \Delta S(GA_TT)^{*3} + \Delta S(AC_TG)^{*3} + \Delta S(CG_GC)^{*3} + \Delta S(\text{loop_3})^{*5} + \Delta S(GGATC)^{*6} = +3.3 - 22.4 - 11.7 - 5.3 - 22.4 - 27.2 - 11.3 + 0 = -97 = -97 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -32.2 - (298.15 \times -97/1000) = -3.279 \text{ kcal mol}^{-1}$$

e) HAIRPINS – case 2

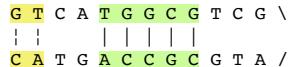


$$\Delta H = \Delta H(AT_XA)^{*3} + \Delta H(TG_AC)^{*3} + \Delta H(GC_CG)^{*3} + \Delta H(CC_GT)^{*3} + \Delta H(CG_TC)^{*3} + \Delta H(GA_CT)^{*3} + \Delta H(AA_TT)^{*3} + \Delta H(\text{loop_4})^{*3} + \Delta H(AG_TG_terminal)^{*7} + \Delta H(AGAAAT)^{*6} = -2.9 - 8.5 - 9.8 - 0.8 - 1.5 - 8.2 - 7.6 + 0 - 1.1 - 2.7 = -43.1 \text{ kcal mol}^{-1}$$

$$\Delta S = \Delta S(AT_XA)^{*3} + \Delta S(TG_AC)^{*3} + \Delta S(GC_CG)^{*3} + \Delta S(CC_GT)^{*3} + \Delta S(CG_TC)^{*3} + \Delta S(GA_CT)^{*3} + \Delta S(AA_TT)^{*3} + \Delta S(\text{loop_4})^{*3} + \Delta S(AG_TG_terminal)^{*7} + \Delta S(AGAAAT)^{*6} = -7.6 - 22.7 - 24.4 - 4.5 - 6.1 - 22.2 - 21.3 - 11.3 - 1.6 - 6.7 = -128.4 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -43.1 - (298.15 \times -128.4/1000) = -4.817 \text{ kcal mol}^{-1}$$

f) HAIRPINS – case 3



$$\Delta H = \Delta H(\text{terminal_AT})^{*2} + \Delta H(TG_AC)^{*3} + \Delta H(GG_CC)^{*3} + \Delta H(GC_CG)^{*3} + \Delta H(CG_GC)^{*3} + \Delta H(\text{loop_6})^{*5} + \Delta H(GT_CG_terminal)^{*7} = +2.2 - 8.5 - 8 - 9.8 - 10.6 + 0 - 4.5 = -39.2 \text{ kcal mol}^{-1}$$

$$\Delta S = \Delta S(\text{terminal_AT})^{*2} + \Delta S(TG_AC)^{*3} + \Delta S(GG_CC)^{*3} + \Delta S(GC_CG)^{*3} + \Delta S(CG_GC)^{*3} + \Delta S(\text{loop_6})^{*5} + \Delta S(GT_CG_terminal)^{*7} = +6.9 - 22.7 - 19.9 - 24.4 - 27.2 - 12.9 - 11.6 = -111.8 = -111.8/1000 = -0.1118 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -39.2 - (298.15 \times -0.1118) = -5.86 \text{ kcal mol}^{-1}$$

$$\Delta H = \Delta H(GT_CA)^{*3} + \Delta H(\text{loop_20})^{*5} + \Delta H(TC_AT_terminal)^{*7} = -8.4 + 0 - 2.7 = -11.1 \text{ kcal mol}^{-1}$$

$$\Delta S = \Delta S(GT_CA)^{*3} + \Delta S(\text{loop_20})^{*5} + \Delta S(TC_AT_terminal)^{*7} = -22.4 - 18.4 - 7.1 = -47.9 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -11.1 - (298.15 \times -47.9/1000) = +3.18 \text{ kcal mol}^{-1}$$

$$\Delta G = \Delta G = -5.86 \text{ kcal mol}^{-1}$$

Figure 11. A schematic view of how the ΔG is calculated for self-dimers (a-b), cross-dimers (c) and hairpin loops (d-e-f). ΔH^{*1} and ΔS^{*1} refer to duplex initiation values; ΔH^{*2} and ΔS^{*2} refer to A and T terminals; ΔH^{*3} and ΔS^{*3} refer to NN base pairs; ΔH^{*4} and ΔS^{*4} refer to symmetry bonus; ΔH^{*5} and ΔS^{*5} refer to loop length corrections; ΔH^{*6} and ΔS^{*6} refer to the 3- and 4- long loop bonus; ΔH^{*7} and ΔS^{*7} refer to loop terminal mismatches. The ΔG was calculated considering a temperature of 25 °C which equals to 298.15 K.

salt-correction formulas that PhyloPrimer applies to the T_m and ΔG calculation (Owczarzy et al., 2004, 2008). Worth mentioning, there are also other elements of a PCR reactions that can influence the melting temperature and the Gibbs energy calculation such as NH_4^+ , DMSO, dyes, glycerol, DMSO, formamide, TMAC and betaine. For these components no correction formulas have been proposed yet. But there are some approximation that the user can look at, such as the one suggested by Ahsen et al., 2001 where it is suggested to decrease the T_m by 0.75 °C for each volume percentage of dimethyl sulfoxide (DMSO).

6.3.1 Monovalent ions and Mg^{2+}

The most used monovalent ions in the PCR reactions are Tris^+ , K^+ and Na^+ . It has been observed that all the monovalent ions act the same for the duplex stabilization. Therefore, they will be all considered as monovalent ions. The concentration of monovalent ions is usually 20-100 mM.

$$[\text{Mon}^+] = [\text{K}^+] + [\text{Tris}^+] + [\text{Na}^+] \quad (3)$$

where:

$[\text{Mon}^+]$ = monovalent ion concentration (M).

$[\text{K}^+]$ = K^+ concentration (M).

$[\text{Tris}^+]$ = Tris^+ concentration (M).

$[\text{Na}^+]$ = Na^+ concentration (M).

It has been observed that Mg^{2+} behavior is more complex than the one observed for monovalent ions. In fact, Mg^{2+} binds stoichiometrically to any of the dNTPs. The bound between Mg^{2+} and dNTP reduces the amount of free Mg^{2+} present for the stabilization of the PCR reaction. Therefore, the first step for understanding the effects of Mg^{2+} is to calculate the concentration of free Mg^{2+} . The concentration of magnesium ions is usually 1.5–5 mM.

If $c_{\text{dNTP}} < 0.8c_{\text{Mg}}$, we assume that the amount of free Mg^{2+} equals the total amount of Mg^{2+} subtracted to the amount of dNTP:

$$[\text{Mg}^{2+}] = c_{\text{Mg}} - c_{\text{dNTP}} \quad (4)$$

If $c_{\text{dNTP}} \geq 0.8c_{\text{Mg}}$, we assume that the amount of free Mg^{2+} is given from the ratio K_a :

$$K_a = \frac{c_{\text{Mg}} - [\text{Mg}^{2+}]}{[\text{Mg}^{2+}](c_{\text{dNTP}} - c_{\text{Mg}} + [\text{Mg}^{2+}])},$$

$$K_a[\text{Mg}^{2+}](c_{\text{dNTP}} - c_{\text{Mg}} + [\text{Mg}^{2+}]) = c_{\text{Mg}} - [\text{Mg}^{2+}],$$

$$K_a([\text{Mg}^{2+}])^2 + (K_a c_{\text{dNTP}} - K_a c_{\text{Mg}} + 1)[\text{Mg}^{2+}] - c_{\text{Mg}} = 0,$$

Resolving the quadratic formula,

$$[\text{Mg}^{2+}] = \frac{-(K_a c_{\text{dNTP}} - K_a c_{\text{Mg}} + 1) \pm \sqrt{(K_a c_{\text{dNTP}} - K_a c_{\text{Mg}} + 1)^2 + 4K_a c_{\text{Mg}}}}{2K_a},$$

As the concentration must be a positive value the square root must be positive, then we obtain:

$$[\text{Mg}^{2+}] = \frac{-(K_a c_{\text{dNTP}} - K_a c_{\text{Mg}} + 1) + \sqrt{(K_a c_{\text{dNTP}} - K_a c_{\text{Mg}} + 1)^2 + 4K_a c_{\text{Mg}}}}{2K_a}. \quad (5)$$

c_{dNTP} = total dNTP concentration (M)

c_{Mg} = total Mg^{2+} concentration (M)

$[\text{Mg}^{2+}]$ = concentration of free Mg^{2+} (M)

K_a = Mg-dNTP association constant, 3×10^{-4} for standard PCR buffers (50 mM KCl and 10 mM Tris)

In order to define which is the dominant ion effect (if the monovalent ions or Mg^{2+}) over the T_m and the secondary structure formations we need to calculate the ratio R:

$$R = \frac{\sqrt{[\text{Mg}^{2+}]}}{[\text{Mon}^+]} \quad (6)$$

6.3.2 Salt-corrected melting temperature

The equation 1 calculates the T_m at 1 M Na^+ condition. In this section we report how to calculate T_m correcting the formula with salt-correction indexes.

If $R < 0.22$, the monovalent ions are considered having a dominant effect over the T_m and the formula for monovalent salt correction is applied:

$$\frac{1}{T_m(\text{Mon}^+)} = \frac{1}{T_m(1 \text{ M Na}^+)} + (4.29f_{GC} - 3.95)10^{-5} \ln[\text{Mon}^+] + 9.40 \times 10^{-6}(\ln[\text{Mon}^+])^2 \quad (7)$$

where:

$T_m(\text{Mon}^+)$ = monovalent-corrected melting temperature (K).

f_{GC} = ratio of Gs/Cs in the oligo.

If $0.22 < R > 6.0$, Mg^{2+} is considered having a dominant effect over the T_m and the formula for Mg^{2+} correction is applied:

$$\frac{1}{T_m(\text{Mg}^{2+})} = \frac{1}{T_m(1 \text{ M Na}^+)} + a + b \ln[\text{Mg}^{2+}] + f_{GC}(c + d \ln[\text{Mg}^{2+}]) + \frac{1}{2(N_{bp} - 1)}[e + f \ln[\text{Mg}^{2+}] + g(\ln[\text{Mg}^{2+}])^2] \quad (8)$$

where:

$T_m(\text{Mg}^{2+})$ = Mg^{2+} -corrected melting temperature (K).

N_{bp} = number of bases in the oligo.

With the parameters a, b, c, d, e, f, g, h and i being:

$$a = 3.92 \times 10^{-5} \quad (9)$$

$$b = -9.11 \times 10^{-6} \quad (10)$$

$$c = 6.26 \times 10^{-5} \quad (11)$$

$$d = 1.42 \times 10^{-5} \quad (12)$$

$$e = -4.82 \times 10^{-4} \quad (13)$$

$$f = 5.25 \times 10^{-4} \quad (14)$$

$$g = 8.31 \times 10^{-5} \quad (15)$$

If $R > 6.0$, the formula for Mg^{2+} correction (equation 8) is applied together with the parameters b, c, e and f which values are reported in the equation 10, 11, 13 and 14, respectively, and the corrected parameters a, d and g :

$$a = 3.92 \times 10^{-5}(0.843 - 0.352\sqrt{[\text{Mon}^+]} \ln[\text{Mon}^+]) \quad (16)$$

$$d = 1.42 \times 10^{-5}[1.279 - 4.03 \times 10^{-3} \ln[\text{Mon}^+] - 8.03 \times 10^{-3}(\ln[\text{Mon}^+])^2] \quad (17)$$

$$g = 8.31 \times 10^{-5}[0.486 - 0.258 \ln[\text{Mon}^+] + 5.25 \times 10^{-3}(\ln[\text{Mon}^+])^3] \quad (18)$$

The forward primer 5'-**ACTCCTCGATCAGGTCGT**-3' has a **T_m of 346.3 K (73.1 °C)** as calculated in Figure 11a and its self-dimer formation shown in Figure 13a has a **ΔH of -37.6 kcal mol⁻¹**, a **ΔS of -110.1 cal K⁻¹ mol⁻¹** and a **ΔG of -4.77 kcal mol⁻¹**. All of these values were calculated assuming that the PCR occurred at 1 M Na⁺.

a)PCR CONDITION – case 1

Mg²⁺ = 0 mM
Oligo = 2 uM = 0.000002 M
dNTP = 2 mM = 0.002 M
Mon+ = 5 mM = 0.005 M

$$R = \sqrt{0}/0.005 = 0 \text{ (eq 6)}$$

As R is lower than 0.22, equation 7 and 19 are applied:

$$\frac{1}{T_m}(\text{Mg}^{2+}) = 1/346.3 + ((4.29*0.56-3.95) \times 10^{-5} \times \ln(0.005)) + (9.4 \times 10^{-6} \times (\ln(0.005)^2)) = 0.0032$$

$$T_m(\text{Mg}^{2+}) = 1/0.0032 = 309.15 \text{ K (36 }^\circ\text{C)}$$

$$\Delta S = -110.1 -37.6 \times ((4.29*0.56-3.95) \times 10^{-5} \times \ln(0.005)) + (9.4 \times 10^{-6} \times (\ln(0.005)^2)) = -110.11235 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -37.6 - (298.15 \times -110.11235/1000) = -4.79 \text{ kcal mol}^{-1}$$

b)PCR CONDITION – case 2

Mg²⁺ = 1.5 mM = 0.0015 M
oligo = 2 uM = 0.000002 M
dNTP = 2 mM = 0.002 M
monovalent ions⁺ = 5 mM = 0.005 M

As 0.002 M > 0.8 x 0.0015 M, equation 5 is applied:

$$\text{Mg}^{2+} = \frac{(-(3 \times 10^4 \times 0.002 - 3 \times 10^4 \times 0.0015 + 1) + \sqrt{(3 \times 10^4 \times 0.002 - 3 \times 10^4 \times 0.0015 + 1)^2 + 4 \times 3 \times 10^4 \times 0.0015})}{2 \times 3 \times 10^4} = 0.00008134 \text{ M (eq 5)}$$

$$R = \sqrt{0.00008134}/0.005 = 1.8 \text{ (eq 6)}$$

As R is > 0.22 and < 6, equation 8 and 20 are applied:

$$\frac{1}{T_m}(\text{Mg}^{2+}) = 1/346.3 + 3.92 \times 10^{-5} - 9.11 \times 10^{-6} \times \ln(0.00008134) + 0.56 \times (6.26 \times 10^{-5} + 1.42 \times 10^{-5} \times \ln(0.00008134)) + (1/2(18 - 1)) \times (-4.82 \times 10^{-4} + 5.25 \times 10^{-4} \times \ln(0.00008134) + 8.31 \times 10^{-5} \times (\ln(0.00008134))^2) = 0.003$$

$$T_m(\text{Mg}^{2+}) = 1/0.00303 = 330 \text{ K (**56.85 }^\circ\text{C**)}$$

$$\Delta S = -110.1 -37.6 \times (3.92 \times 10^{-5} - 9.11 \times 10^{-6} \times \ln(0.00008134) + 0.56 \times (6.26 \times 10^{-5} + 1.42 \times 10^{-5} \times \ln(0.00008134)) + (1/2(10 - 1)) \times (-4.82 \times 10^{-4} + 5.25 \times 10^{-4} \times \ln(0.00008134) + 8.31 \times 10^{-5} \times (\ln(0.00008134))^2)) = -110.10665 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -37.6 - (298.15 \times -110.10665/1000) = -4.78 \text{ kcal mol}^{-1}$$

c)PCR CONDITION – case 3

Mg²⁺ = 5 mM = 0.005 M
oligo = 2 uM = 0.000002 M
dNTP = 1 mM = 0.002 M
monovalent ions⁺ = 5 mM = 0.005 M

As 0.002 M is < 0.8 x 0.005 M, we apply (4) for calculating the free Mg²⁺

$$\text{Mg}^{2+} = 0.005 - 0.002 = 0.003 \text{ M (eq 4)}$$

$$R = \sqrt{0.003}/0.005 = 10.95 \text{ (eq 6)}$$

As R is higher than 6, a, d and g are corrected with equation 16, 17 and 18:

$$a = 3.82 \times 10^{-5}$$

$$d = 1.52 \times 10^{-5}$$

$$g = 8.90 \times 10^{-5}$$

And equation 8 and 20 are applied:

$$\frac{1}{T_m}(\text{Mg}^{2+}) = 1/346.3 + 3.82 \times 10^{-5} - 9.11 \times 10^{-6} \times \ln(0.003) + 0.56 \times (6.26 \times 10^{-5} + 1.52 \times 10^{-5} \times \ln(0.003)) + (1/2(18 - 1)) \times (-4.82 \times 10^{-4} + 5.25 \times 10^{-4} \times \ln(0.003) + 8.9 \times 10^{-5} \times (\ln(0.003))^2) = 0.00294$$

$$T_m(\text{Mg}^{2+}) = 1/0.00294 = 339.11 \text{ K (**65.96 }^\circ\text{C**)}$$

$$\Delta S = -110.1 -37.6 \times (3.82 \times 10^{-5} - 9.11 \times 10^{-6} \times \ln(0.003) + 0.56 \times (6.26 \times 10^{-5} + 1.52 \times 10^{-5} \times \ln(0.003)) + (1/2(10 - 1)) \times (-4.82 \times 10^{-4} + 5.25 \times 10^{-4} \times \ln(0.003) + 8.9 \times 10^{-5} \times (\ln(0.003))^2)) = -110.10181 \text{ cal K}^{-1} \text{ mol}^{-1}$$

$$\Delta G = -37.6 - (298.15 \times -110.10181/1000) = -4.79 \text{ kcal mol}^{-1}$$

Figure 12. A schematic view of how the T_m and ΔG values are calculated at different salt-conditions. The ΔG was calculated considering a temperature of 25 °C which equals to 298.15 K.

In Figure 12 we show an example of how the melting temperature was corrected for different PCR conditions. We can observe how the temperature changed drastically. The melting temperature calculated without any salt-correction was 73.1 °C (Figure 10a), when the monovalent ions had a dominant effect ($R < 0.22$; Figure 12a) the T_m was 36 °C, whereas when the R was in between 0.22 and 6, the T_m was 56.85 °C (Figure 12b) and with a $R > 6.0$ the T_m was 65.96 °C (Figure 12c).

6.3.3 Salt-corrected ΔG

That equation 2 was for obtaining the ΔG in 1 M Na^+ conditions. Whereas ΔH is considered salt-independent, ΔS varies with different PCR conditions. We, therefore, need to correct ΔS in relation to the salt conditions.

$$\frac{\Delta S^\circ(\text{Mon}^+) + R \ln(\frac{C_T}{x})}{\Delta H^\circ} = \frac{\Delta S^\circ(1 \text{ M } \text{Na}^+) + R \ln(\frac{C_T}{x})}{\Delta H^\circ} + (4.29f_{GC} - 3.95)10^{-5} \ln[\text{Mon}^+] + 9.40 \times 10^{-6} (\ln[\text{Mon}^+])^2,$$

$$\Delta S^\circ(\text{Mon}^+) + R \ln(\frac{C_T}{x}) = \Delta S^\circ(1 \text{ M } \text{Na}^+) + R \ln(\frac{C_T}{x}) + \Delta H^\circ \{(4.29f_{GC} - 3.95)10^{-5} \ln[\text{Mon}^+] + 9.40 \times 10^{-6} (\ln[\text{Mon}^+])^2\},$$

$$\Delta S^\circ(\text{Mon}^+) = \Delta S^\circ(1 \text{ M } \text{Na}^+) + \Delta H^\circ \{(4.29f_{GC} - 3.95)10^{-5} \ln[\text{Mon}^+] + 9.40 \times 10^{-6} (\ln[\text{Mon}^+])^2\}. \quad (19)$$

If $0.22 < R > 6.0$, Mg^{2+} is considered having a dominant effect over the ΔS and the formula for Mg^{2+} correction is applied. We can substitute the equation 8 with the the equation 2, which describes how to calculate ΔG ,

$$\begin{aligned} \frac{\Delta S^\circ(\text{Mg}^{2+}) + R \ln(\frac{C_T}{x})}{\Delta H^\circ} &= \\ \frac{\Delta S^\circ(1 \text{ M } \text{Na}^+) + R \ln(\frac{C_T}{x})}{\Delta H^\circ} &+ a + b \ln[\text{Mg}^{2+}] + f_{GC}(c + d \ln[\text{Mg}^{2+}]) + \frac{1}{2(N_{bp} - 1)} [e + f \ln[\text{Mg}^{2+}] + g(\ln[\text{Mg}^{2+}])^2], \\ \Delta S^\circ(\text{Mg}^{2+}) + R \ln(\frac{C_T}{x}) &= \\ \Delta S^\circ(1 \text{ M } \text{Na}^+) + R \ln(\frac{C_T}{x}) &+ \Delta H^\circ \{a + b \ln[\text{Mg}^{2+}] + f_{GC}(c + d \ln[\text{Mg}^{2+}]) + \frac{1}{2(N_{bp} - 1)} [e + f \ln[\text{Mg}^{2+}] + g(\ln[\text{Mg}^{2+}])^2]\}, \\ \Delta S^\circ(\text{Mg}^{2+}) &= \\ \Delta S^\circ(1 \text{ M } \text{Na}^+) + \Delta H^\circ \{a + b \ln[\text{Mg}^{2+}] &+ f_{GC}(c + d \ln[\text{Mg}^{2+}]) + \frac{1}{2(N_{bp} - 1)} [e + f \ln[\text{Mg}^{2+}] + g(\ln[\text{Mg}^{2+}])^2]\}. \end{aligned} \quad (20)$$

If $R > 6.0$, the formula for Mg^{2+} correction (equation 20) is applied together with the parameters b , c , e and f reported in the equations 10, 11, 12 and 13, respectively, and the corrected parameters a , d and g from the equations 16, 17 and 18, respectively.

Once the salt-corrected ΔS is obtained, ΔG can be calculated with the equation 2. In the Figure 12 we can observe how the ΔG value is much less susceptible to change due to the PCR conditions than the T_m value. The ΔG calculated for the self-dimer secondary structure in Figure 11a was -4.77 kcal mol⁻¹, whereas with the different salt-correction formulas applied in Figure 12, that values ranged between -4.78 and -4.79 kcal mol⁻¹.

7 Database

PhyloPrimer uses external nucleotide sequence databases in two points of the pipeline. The first point is when it BLAST searches the sequences uploaded in the Dynamic Selection mode to retrieve similar sequences and construct a dynamic phylogenetic tree (DB1), and the second when it checks the oligo specificity through in silico

PCR (DB2). DB1 is constituted by protein, rRNA, tRNA and tmRNA coding regions annotated from GenBank prokaryotic genomes (Sayers et al., 2019). Nucleotide sequences from a maximum of 50 different genome assemblies or complete genomes are reported per organism for a total of 70,812 bacterial genomes and 3,007 archaeal genomes. DB2 is the nt database (<ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nt.gz>) which contains partially non-redundant nucleotide sequences from the GenBank, EMBL and DDBJ databases. The two databases were last updated in October 2020 and are updated every two months. At the moment of the publication they contained 261,143,432 (DB1) and 60,553,916 (DB2) entries.

8 Server Specifics and Coding language

The current server is a 64-bit Intel(R) Xeon(R) CPU E5-2680 v3 at 2.50GHz (64 CPUs). Only 6 PhyloPrimer processes at one time are allowed on the server, the excess processes will enter a queue. On average, the oligo design requires 30/40 minutes whereas the oligo check requires 5/10 minutes. PhyloPrimer is coded in Perl, JavaScript, HTML, CSS and MySQL. Two javascript packages were used: a modified version of PhyloCanvas (<http://phylocanvas.org>) and canvasJS (<https://canvasjs.com>). All the PhyloPrimer scripts are also available through the HitHUB page: <https://github.com/gvMicroarctic/PhyloPrimer>.

Appendices

Table A1. ΔH and ΔS values for NN base pairs (a), internal mismatches (b) and dangling ends (c).

| a | NN pairs | ΔH (kcal mol ⁻¹) | ΔS (cal K ⁻¹ mol ⁻¹) | b | NN pairs | ΔH (kcal mol ⁻¹) | ΔS (cal K ⁻¹ mol ⁻¹) |
|--------------|---------------|---|--|---|-----------|---|--|
| | AA_TT | -7.6 | -21.3 | | AA_TA | 1.2 | 1.7 |
| | AT_TA | -7.2 | -20.4 | | CA_GA | -0.9 | -4.2 |
| | TA_AT | -7.2 | -21.3 | | GA_CA | -2.9 | -9.8 |
| | CA_GT | -8.5 | -22.7 | | TA_AA | 4.7 | 12.9 |
| | GT_CA | -8.4 | -22.4 | | AC_TC | 0.0 | -4.4 |
| | CT_GA | -7.8 | -21.0 | | CC_GC | -1.5 | -7.2 |
| | GA_CT | -8.2 | -22.2 | | GC_CC | 3.6 | 8.9 |
| | CG_GC | -10.6 | -27.2 | | TC_AC | 6.1 | 16.4 |
| | GG_CC | -8.0 | -19.9 | | AG_TG | -3.1 | -9.5 |
| initiation | 0.2 | -5.7 | | | CG_GG | -4.9 | -15.3 |
| terminal A.T | 2.2 | 6.9 | | | GG(CG) | -6.0 | -15.8 |
| symmetry | 0.0 | -1.4 | | | TG_AG | 1.6 | 3.6 |
| | | | | | AT_TT | -2.7 | -10.8 |
| c | Dangling ends | ΔH (kcal mol ⁻¹) | ΔS (cal K ⁻¹ mol ⁻¹) | | CT_GT | -5.0 | -15.8 |
| | | | | | GT_CT | -2.2 | -8.4 |
| | AA_XT | 0.2 | 2.3 | | TT_AT | 0.2 | -1.5 |
| | AC_XG | -6.3 | -17.1 | | AG_TT | 1.0 | 0.9 |
| | AG_XC | -3.7 | -10.0 | | AT_TG | -2.5 | -8.3 |
| | AT_XA | -2.9 | -7.6 | | CG_GT | -4.1 | -11.7 |
| | CA_XT | 0.6 | 3.3 | | CT_GG | -2.8 | -8.0 |
| | CC_XG | -4.4 | -12.6 | | GG_CT | 3.3 | 10.4 |
| | CG_XC | -4.0 | -11.9 | | GT(CG)_CG | -4.4 | -12.3 |
| | CT_XA | -4.1 | -13.0 | | TG_AT | -0.1 | -1.7 |
| | GA_XT | -1.1 | -1.6 | | TT_AG | -1.3 | -5.3 |
| | GC_XG | -5.1 | -14.0 | | AA_TC | 2.3 | 4.6 |
| | GG_XC | -3.9 | -10.9 | | AC_TA | 5.3 | 14.6 |
| | GT_XA | -4.2 | -15.0 | | CA_GC | 1.9 | 3.7 |
| | TA_XT | -6.9 | -20.0 | | CC_GA | 0.6 | -0.6 |
| | TC_XG | -4.0 | -10.9 | | GA_CC | 5.2 | 14.2 |
| | TG_XC | -4.9 | -13.8 | | GC_CA | -0.7 | -3.8 |
| | TT_XA | -0.2 | -0.5 | | TA_AC | 3.4 | 8.0 |
| | XA_AT | -0.7 | -0.8 | | TC_AA | 7.6 | 20.2 |
| | XC_AG | -2.1 | -3.9 | | AA_TG | -0.6 | -2.3 |
| | XG_AC | -5.9 | -16.5 | | AG_TA | -0.7 | -2.3 |
| | XT_AA | -0.5 | -1.1 | | CA_GG | -0.7 | -2.3 |
| | XA_CT | 4.4 | 14.9 | | CG_GA | -4.0 | -13.2 |
| | XC_CG | -0.2 | -0.1 | | GA(CG)_CG | -0.6 | -1.0 |
| | XG_CC | -2.6 | -7.4 | | GG_CA | 0.5 | 3.2 |
| | XT_CA | 4.7 | 14.2 | | TA_AG | 0.7 | 0.7 |
| | XA_GT | -1.6 | -3.6 | | TG_AA | 3.0 | 7.4 |
| | XC_GG | -3.9 | -11.2 | | AC_TT | 0.7 | 0.2 |
| | XG_GC | -3.2 | -10.4 | | AT_TC | -1.2 | -6.2 |
| | XT_GA | -4.1 | -13.1 | | CC_GT | -0.8 | -4.5 |
| | XA_TT | 2.9 | 10.4 | | CT_GC | -1.5 | -6.1 |
| | XC_TG | -4.4 | -13.1 | | GC_CT | 2.3 | 5.4 |
| | XG_TC | -5.2 | -15.0 | | GT_CC | 5.2 | 13.5 |
| | XT_TA | -3.8 | -12.6 | | TC_AT | 1.2 | 0.7 |
| | | | | | TT_AC | 1.0 | 0.7 |

Table A2. ΔH and ΔS values for loop-length correction (A), tri-loop bonus (B) and tetra-loop bonus (C).

| a | Loop length | ΔH (kcal mol ⁻¹) | ΔS (cal K ⁻¹ mol ⁻¹) | c | Tetra-loops | ΔH (kcal mol ⁻¹) | ΔS (cal K ⁻¹ mol ⁻¹) | Tetra-loops | ΔH (kcal mol ⁻¹) | ΔS (cal K ⁻¹ mol ⁻¹) |
|---|-------------|---|--|---|-------------|---|--|-------------|---|--|
| | 3 | 0.0 | -11.3 | | AAAAAT | 0.5 | -0.6 | GAAAAT | 0.5 | -3.2 |
| | 4 | 0.0 | -11.3 | | AAACT | 0.7 | 1.6 | GAAACT | 1.0 | 0.0 |
| | 5 | 0.0 | -10.6 | | AAACAT | 1.0 | 1.6 | GAACAT | 1.0 | 0.0 |
| | 6 | 0.0 | -12.9 | | ACTTGT | 0.0 | 4.2 | GCTTGT | 0.0 | 1.6 |
| | 7 | 0.0 | -13.5 | | AGAAAT | -1.1 | 1.6 | GGAAAT | -1.1 | 0.0 |
| | 8 | 0.0 | -13.9 | | AGAGAT | -1.1 | 1.6 | GGAGAT | -1.1 | 0.0 |
| | 9 | 0.0 | -14.5 | | AGATAT | -1.5 | 1.6 | GGATAT | -1.6 | 0.0 |
| | 10 | 0.0 | -14.8 | | AGCAAT | -1.6 | 1.6 | GGCAAT | -1.6 | 0.0 |
| | 11 | 0.0 | -15.5 | | AGCGAT | -1.1 | 1.6 | GGCGAT | -1.1 | 0.0 |
| | 12 | 0.0 | -16.1 | | AGCTTT | 0.2 | 1.6 | GGCTTT | -0.1 | 0.0 |
| | 13 | 0.0 | -16.4 | | AGGAAT | -1.1 | 1.6 | GGGAAT | -1.1 | 0.0 |
| | 14 | 0.0 | -16.4 | | AGGGAT | -1.1 | 1.6 | GGGGAT | -1.1 | 0.0 |
| | 15 | 0.0 | -16.8 | | AGGGGT | 0.5 | 0.6 | GGGGGT | 0.5 | -1.0 |
| | 16 | 0.0 | -17.1 | | AGTAAT | -1.6 | 1.6 | GGTAAT | -1.6 | 0.0 |
| | 17 | 0.0 | -17.4 | | AGTGAT | -1.1 | 1.6 | GGTGAT | -1.1 | 0.0 |
| | 18 | 0.0 | -17.7 | | AGTTCT | 0.8 | 1.6 | GTATAT | -0.5 | 0.0 |
| | 19 | 0.0 | -18.1 | | ATTCGT | -0.2 | 1.6 | GTTCGT | -0.4 | 0.0 |
| | 20 | 0.0 | -18.4 | | ATTGTT | 0.0 | 1.6 | GTGGAT | -0.4 | 0.0 |
| | | | | | ATTTTT | -0.5 | 1.6 | GTTTTT | -0.5 | 0.0 |
| b | Tri-loops | ΔH (kcal mol ⁻¹) | ΔS (cal K ⁻¹ mol ⁻¹) | | CAAAAG | 0.5 | -1.3 | TAAAAA | 0.5 | 0.3 |
| | | | | | CAAACG | 0.7 | 0.0 | TAAACA | 0.7 | 1.6 |
| | AGAAT | -1.5 | 0.0 | | CAACAG | 1.0 | 0.0 | TAACAA | 1.0 | 1.6 |
| | AGCAT | -1.5 | 0.0 | | CAACCG | 0.0 | 0.0 | TCTTGA | 0.0 | 4.2 |
| | AGGAT | -1.5 | 0.0 | | CCTTGG | 0.0 | 2.6 | TGAAAA | -1.1 | 1.6 |
| | AGTAT | -1.5 | 0.0 | | CGAAAG | -1.1 | 0.0 | TGAGAA | -1.1 | 1.6 |
| | CGAAC | -2.0 | 0.0 | | CGAGAG | -1.1 | 0.0 | TGATAA | -1.6 | 1.6 |
| | CGCAG | -2.0 | 0.0 | | CGATAG | -1.5 | 0.0 | TGCAAA | -1.6 | 1.6 |
| | CGGAG | -2.0 | 0.0 | | CGCAAG | -1.6 | 0.0 | TGCGAA | -1.1 | 1.6 |
| | CGTAG | -2.0 | 0.0 | | CGCGAG | -1.1 | 0.0 | TGCTTA | 0.2 | 1.6 |
| | GGAAC | -2.0 | 0.0 | | CGCTTG | 0.2 | 0.0 | TGGAAA | -1.1 | 1.6 |
| | GGCAC | -2.0 | 0.0 | | CGGAAG | -1.1 | 0.0 | TGGGAA | -1.1 | 1.6 |
| | GGGAC | -2.0 | 0.0 | | CGGGAG | -1.0 | 0.0 | TGGGGA | 0.5 | 0.6 |
| | GGTAC | -2.0 | 0.0 | | CGGGGG | 0.5 | -1.0 | TGTAAA | -1.6 | 1.6 |
| | TGAAA | -1.5 | 0.0 | | CGTAAG | -1.6 | 0.0 | TGTGAA | -1.1 | 1.6 |
| | TGCAA | -1.5 | 0.0 | | CGTGAG | -1.1 | 0.0 | TGTTCA | 0.8 | 1.6 |
| | TGGAA | -1.5 | 0.0 | | CGTTCG | 0.8 | 0.0 | TTTCGA | -0.2 | 1.6 |
| | TGTAA | -1.5 | 0.0 | | CTTCGG | -0.2 | 0.0 | TTTTGA | 0.0 | 1.6 |
| | | | | | CTTTGG | 0.0 | 0.0 | TTTTTA | -0.5 | 1.6 |
| | | | | | CTTTTG | -0.5 | 0.0 | TAAAAG | 0.5 | -1.6 |
| | | | | | GAAAAC | 0.5 | -3.2 | TAAACG | 1.0 | 1.6 |
| | | | | | GAAACC | 0.7 | 0.0 | TAACAG | 1.0 | 1.6 |
| | | | | | GAACAC | 1.0 | 0.0 | TCTTGG | 0.0 | 3.2 |
| | | | | | GCTTGC | 0.0 | 2.6 | TGAAAG | -1.0 | 1.6 |
| | | | | | GGAAAC | -1.1 | 0.0 | TGAGAG | -1.0 | 1.6 |
| | | | | | GGAGAC | -1.1 | 0.0 | TGATAG | -1.5 | 1.6 |
| | | | | | GGATAC | -1.6 | 0.0 | TGCAAG | -1.5 | 1.6 |
| | | | | | GGCAAC | -1.6 | 0.0 | TGCGAG | -1.0 | 1.6 |
| | | | | | GGCGAC | -1.1 | 0.0 | TGCTTG | -0.1 | 1.6 |
| | | | | | GGCTTC | 0.2 | 0.0 | TGGAAG | -1.0 | 1.6 |
| | | | | | GGGAAC | -1.1 | 0.0 | TGGGAG | -1.0 | 1.6 |
| | | | | | GGGGAC | -1.1 | 0.0 | TGGGGG | 0.5 | 0.6 |
| | | | | | GGGGC | 0.5 | -1.0 | TGTAAG | -1.5 | 1.6 |
| | | | | | GCTAAC | -1.6 | 0.0 | TGTGAG | -1.0 | 1.6 |
| | | | | | GGTGAC | -1.1 | 0.0 | TTTCGG | -0.4 | 1.6 |
| | | | | | GGTTCC | 0.8 | 0.0 | TTTTAG | -1.0 | 1.6 |
| | | | | | GTTCGC | -0.2 | 0.0 | TTTTGG | -0.4 | 1.6 |
| | | | | | GTTTGC | 0.0 | 0.0 | TTTTTG | -0.5 | 1.6 |
| | | | | | GTTTTC | -0.5 | 0.0 | | | |

Table A3. ΔH and ΔS values for loop terminal mismatches.

| NN pairs | ΔH (kcal mol $^{-1}$) | ΔS (cal K $^{-1}$ mol $^{-1}$) | NN pairs | ΔH (kcal mol $^{-1}$) | ΔS (cal K $^{-1}$ mol $^{-1}$) |
|-----------|-----------------------------------|--|----------|-----------------------------------|--|
| AA_TA | -3.2 | -8.1 | GC_TA | 0.0 | 0.6 |
| AA_TC | -0.9 | -1.9 | GC_TC | 0.0 | 0.6 |
| AA_TG | -2.3 | -5.8 | GC_TG | -4.5 | -11.6 |
| AC_TA | -2.2 | -5.2 | GC_TT | 0.0 | 0.6 |
| AC_TC | -0.5 | -1.0 | GG_TA | 0.0 | 1.6 |
| AC_TT | -1.2 | -2.9 | GG_TC | -5.9 | -16.1 |
| AG_TA | -2.7 | -6.8 | GG_TG | 0.0 | 1.6 |
| AG_TG | -1.3 | -2.9 | GG_TT | -2.0 | -4.8 |
| AG_TT | -2.9 | -7.7 | GT_TA | -3.5 | -9.7 |
| AT_TC | -2.8 | -8.1 | GT_TC | 0.0 | 0.6 |
| AT_TG | -3.5 | -9.7 | GT_TG | -2.0 | -4.8 |
| AT_TT | -2.4 | -6.4 | GT_TT | 0.0 | 0.6 |
| CA_GA | -2.8 | -5.8 | TA_AA | -2.7 | -6.8 |
| CA_GC | -2.0 | -3.9 | TA_AC | -2.6 | -7.1 |
| CA_GG | -3.0 | -6.8 | TA_AG | -2.4 | -6.1 |
| CC_GA | -2.4 | -5.2 | TC_AA | -2.6 | -6.8 |
| CC_GC | -1.4 | -2.9 | TC_AC | -0.5 | -1.0 |
| CC_GT | -2.4 | -5.5 | TC_AT | -2.7 | -7.1 |
| CG_GA | -5.1 | -13.2 | TG_AA | -1.9 | -4.2 |
| CG_GG | -2.9 | -6.4 | TG_AG | -1.5 | -3.5 |
| CG_GT | -2.9 | -6.1 | TG_AT | -2.3 | -5.8 |
| CT_GC | -3.1 | -8.1 | TT_AC | -1.4 | -3.5 |
| CT_GG | -5.9 | -16.1 | TT_AG | -3.7 | -10 |
| CT_GT | -5.3 | -14.2 | TT_AT | -2.3 | -6.4 |
| GA_CA | -6.0 | -16.1 | TA_GA | 0.0 | 1.6 |
| GA_CC | -4.0 | -10.6 | TA_GC | 0.0 | 0.6 |
| GA_CG | -3.4 | -8.4 | TA_GG | 0.0 | 1.6 |
| GC_CA | -2.8 | -5.8 | TA_GT | -2.3 | -5.8 |
| GC_CC | -2.5 | -6.1 | TC_GA | 0.0 | 0.6 |
| GC_CT | -3.3 | -8.4 | TC_GC | 0.0 | 0.6 |
| GG_CA | -4.0 | -9.7 | TC_GG | -3.7 | -9.4 |
| GG(CG)_CG | -5.3 | -13.9 | TC_GT | 0.0 | 0.6 |
| GG_CT | -3.7 | -9.4 | TG_GA | 0.0 | 1.6 |
| GT_CC | -2.5 | -6.1 | TG_GC | -2.9 | -6.1 |
| GT(CG)_CG | -4.5 | -11.6 | TG_GG | 0.0 | 1.6 |
| GT_CT | -6.1 | -16.8 | TG_GT | -2.0 | -4.8 |
| GA_TA | 0.0 | 1.6 | TT_GA | -2.9 | -7.7 |
| GA_TC | 0.0 | 0.6 | TT_GC | 0.0 | 0.6 |
| GA_TG | 0.0 | 1.6 | TT_GG | -2.0 | -4.8 |
| GA_TT | -3.7 | -10.0 | TT_GT | 0.0 | 0.6 |