TECHNICAL NOTE

OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes

Marc Lohse · Oliver Drechsel · Ralph Bock

Abstract Mitochondria and plastids are DNA-containing cell organelles whose genomes occur at high copy numbers per cell. Organellar genomes vary greatly in size ranging from approximately 15 kb for some animal mitochondrial genomes to more than 2 Mb for some plant mitochondrial genomes. The vast majority of organellar genomes map as circular molecules that are difficult to illustrate by available commercial or free software tools. Thus, published genome maps are extremely heterogeneous in design, often tediously drawn semi-manually and lack any consensus in display. Here, we present a new web-based tool, OrganellarGenomeDRAW (OGDRAW), which produces high-resolution custom graphical maps of DNA sequences as stored in standard GenBank format entries. GenBank data can be provided as either file uploads or accession numbers. The program is specially optimized for the display of chloroplast and mitochondrial genomes but can also be used to depict other circular DNA sequences. The design of the program core as a Perl module with an object-oriented interface allows easy integration into custom scripts.

Keywords Plastid · Mitochondrion · Physical map · Restriction map · Bioinformatics · Software tool

Introduction

The genetic information of eukaryotes is distributed among either two or three cellular compartments. In animal and fungal cells, it is contained in the nucleus and the mitochondria, whereas plant cells harbor an additional DNA-containing organelle, the plastid (chloroplast). Mitochondria and plastids are of endosymbiotic origin and were acquired by a pre-eukaryotic host cell through engulfment of free-living α-proteobacteria and cyanobacteria, respectively. Consistent with their prokaryotic history, the gene expression systems of present-day mitochondria and plastids still exhibit numerous eubacterial features, including gene organization in operons that are co-transcribed as polycistronic mRNAs and translated on prokaryotic-type 70S ribosomes.

Plastid genomes (ptDNA) and most mitochondrial genomes (mtDNA) map as circular double-stranded DNA molecules. Compared to the nuclear genome, the genomes of plastids and mitochondria contain comparably little information. The ptDNA of most land plants shows a tetrapartite genome organization with a large single copy region (LSC) and a small single copy region (SSC) separating two inverted repeat regions (IR_A and IR_B). The two IRs are identical in their nucleotide sequence and differ only in their relative orientation. The ptDNA of most land plant species harbors a rather conserved set of approximately 100–130 genes in a genome of 120–160 kb (reviewed, e.g., in Wakasugi et al. 2001; Bock 2007). Exceptions include holoparasitic plants which often possess much smaller plastid genomes due to the deletion of photosynthesis genes that are no longer needed after having switched to a parasitic lifestyle (dePamphilis and Palmer 1990; Bungard 2004). *Pelargonium*, the genus of the flowering plants with the largest ptDNA, has a genome size of 217 kb (Palmer et al. 1987; Chumley et al. 2006). This increase in genome size is, however, not due to the presence of additional genes, but instead can be chiefly attributed to a large expansion of the inverted repeats. In contrast to higher plants,

M. Lohse · O. Drechsel · R. Bock (✉)
Max-Planck-Institut für Molekulare Pflanzenphysiologie,
Am Mühlenberg 1, 14476 Potsdam-Golm, Germany
e-mail: rbock@mpimp-golm.mpg.de

ptDNA size and coding capacity in algae are much more variable. In some algal lineages, the plastid genomes are extremely compact and gene-dense. For example, the ptDNA of the cryptophyte alga *Guillardia theta* harbors as many as 180 genes in a genome of only 122 kb (Douglas and Penny 1999). In other algae, the ptDNA has greatly expanded in size, mainly by the accumulation of non-coding DNA. Such a genome expansion has occurred in the model organism *Chlamydomonas reinhardtii*, a unicellular green alga, whose ptDNA is 204 kb large, but contains only 99 genes (Maul et al. 2002).

Higher plant mitochondrial genomes have an even lower coding capacity than plastid genomes and typically harbor only approximately 60 genes (Unseld et al. 1997; Knoop 2004). Nonetheless, they are often bigger than plastid genomes and moreover, display great size variation between species, ranging from 180 kb to 2.4 Mb. Thus, higher plant mtDNAs are also much larger and more complex in their genome organization than the mitochondrial genomes of protists (5.7–76 kb), animals (14–42 kb) and fungi (18–176 kb; Backert et al. 1997; Gray et al. 1998). The enormous size differences between the highly compact mtDNAs of animals and the large mtDNAs of higher plants are mainly due to the presence of large non-coding intergenic spacers, introns and duplicated sequences in plant mitochondrial genomes (Unseld et al. 1997; Backert et al. 1997; Knoop 2004).

The first fully sequenced genomes from living organisms (excluding phages) were organellar genomes. The complete sequence of the human mitochondrial genome was determined in 1981 (Anderson et al. 1981) and 5 years later, the first two fully sequenced plastid genomes followed (Ohyama et al. 1986; Shinozaki et al. 1986). With the rapidly improving sequencing technologies, the past few years have witnessed an explosion in the sequencing of complete organellar genomes, and to date, more than a thousand mitochondrial genomes and more than a hundred plastid genomes have been fully sequenced (http://www.ncbi.nlm.nih.gov/genomes/static/euk_o.html) representing all major lineages of eukaryotic evolution. All this sequence information is publicly available and stored in the form of GenBank entries. Typically, these entries not only provide the raw nucleotide sequence but also annotation information describing the gene loci and other sequence features that were experimentally determined or bioinformatically predicted.

Several software tools for molecular biologists offer the option to create graphical maps of DNA sequences. General molecular biology software packages, like Vector NTI (Lu and Moriyama 2004) or DNAStar (Burland 2000), and some freeware tools (e.g., PlasMapper; Dong et al. 2004) are able to read GenBank or FASTA files, but were mainly designed to display plasmid maps and, unfortunately, fail to generate clearly structured maps of larger sequences (PlasMapper, for example, actually rejects sequences larger than 20 kb). More specialized tools (e.g., CGView: Stothard and Wishart 2005; Tsudzuki 2000) can produce decent maps of some organellar genome sequences, but offer little convenience and flexibility and, moreover, require special input file formats that must be generated manually.

Since, to date, there is no convenient and user-friendly software tool available to directly visualize circular genomes as clearly laid out, high-quality graphical maps, we have designed OrganellarGenomeDRAW (OGDRAW). OGDRAW is available online, extremely simple to use and offers a variety of versatile features allowing the user to draw custom-made, publishable-quality genome maps from GenBank entries.

Results and discussion

Program description

The OrganellarGenomeDRAW tool is structured into two parts. A graphical user interface (Fig. 1) that is dynamically generated by a Perl/CGI script allows the user to conveniently control most of the display options. A back-end (written mainly in Perl) then processes the supplied information and generates the output file. The direct output of the program is a PostScript© file that is converted to the user-chosen graphics format using the free ImageMagick (http://www.imagemagick.org/) graphics library.

To generate a map, the program extracts the raw annotation information for each sequence feature stored in the provided GenBank file and processes it in a multistep procedure: Duplicated entries and entries that have identical start and end positions but different labels are filtered and shown as one feature in the map. Very large features (e.g., genes that are trans-spliced) will be dissected into their sublocations, and only the sublocations will be drawn in the map. This step is necessary, because in GenBank entries, trans-spliced genes are normally annotated as one large feature and thus could easily cover half of the genome map. For example, the trans-spliced exons of the *rps12* gene in higher plant plastid genomes are more than 70 kb apart and, if the *rps12* feature in the GenBank entry would not be processed by the dissection procedure, *rps12* would be drawn as one giant gene covering more than 70 kb. The dissection process built into the program facilitates the separate display of the trans-spliced gene pieces, thereby preventing interference with the display of all the many genes in between the trans-spliced *rps12* exons.

A human-readable name is extracted for each feature in the GenBank entry and, together with the type of the sequence feature (protein-coding sequence, tRNA, etc.), is

Fig. 1 Screenshot of the web interface of OGDRAW showing the main features that can be selected for display in the graphical genome map. This screenshot shows the options selected for drawing the map in Fig. 2

used to determine the class the feature belongs to. Typical organellar genomes contain a conserved set of genes, most of which encode subunits of a limited number of major protein complexes (Figs. 2, 3). These include protein complexes involved in gene expression (e.g., ribosome, RNA polymerase) and the large multiprotein complexes present in energy-transducing membranes: the protein complexes of the respiratory chain in mitochondria, and the complexes of the photosynthetic electron transport chain in plastids. To improve the clarity and comparability of the produced maps and to help establishing a consensus color coding of organellar genome maps, all genes encoding subunits of these protein complexes will be automatically classified (by text searching of the GenBank annotations) and color coded making their complex affiliation visible at a glance. Different sets of default class definitions (corresponding to the major complexes that are at least in part encoded by the plastid and mitochondrial genomes, respectively) are integrated in the program. Furthermore, the user can take complete control over the feature class definitions by providing a custom configuration file (Box 1). The configuration can be supplied in a simple XML-like format that defines a set

of parameters for each feature class, including its name, type and the color it should be drawn in (Box 1; Fig. 1). To use a custom configuration, the user just has to generate the file (e.g., by modifying the default file that is available for download in the FAQ section of the OGDRAW web site) and specify the file name in the submission form of the web site. Operons can additionally be displayed in the map as polycistronic transcription units, but this requires that the corresponding primary transcripts are properly annotated as 'prim_transcript'-type features in the GenBank entry. If the 'tidy-up' option is activated (Fig. 1), the program will also eliminate from the map single features that are larger than 10 kb and could not be dissected into sublocations. It will also try to reformat tRNA and rRNA gene names according to the naming conventions detailed below. For quick and precise detection of inverted repeat regions in organellar genomes, a small D program (http://www.digitalmars.com/d/index.html) was developed and included in the package.

After finishing the preprocessing, the curated list of sequence features is processed to a PostScript© output file that can either be obtained directly or, as the final step, can be converted to the graphics format chosen by the user
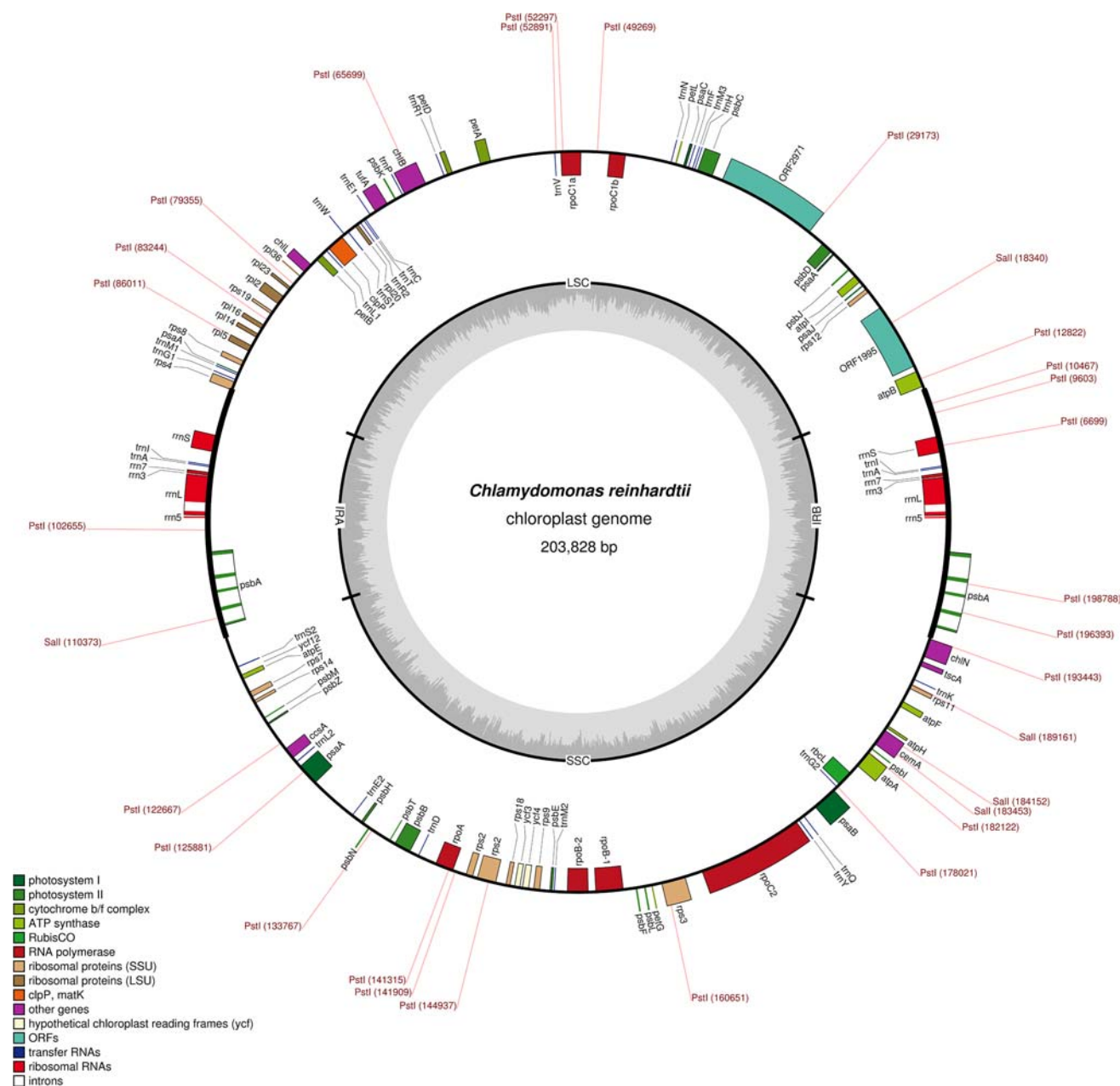
Fig. 2 Physical map of the plastid genome of the unicellular green alga *Chlamydomonas reinhardtii* (BK000554.2) as drawn by OG-DRAW. Genes *inside of the circle* are transcribed clockwise, genes *outside the circle* are transcribed counterclockwise. The 'automatically detect IR borders' function was activated. Activation of the 'tidy-up' function is not necessary (see Fig. 1 for a list of the selected options). For comparison, see the previously published low-quality homemade physical map (Maul et al. 2002). Additional features that were selected

here for display by OGDRAW include the GC content graph (*inner ring*) and the restriction sites for the endonucleases PstI and SalI. The unusual naming of the tRNA and rRNA genes in the original GenBank entry can be changed to the standard nomenclature by simply downloading the GenBank entry and editing the gene names in a text editor. OGDRAW can then use the modified file for displaying the custom-edited map

(Figs. 2, 3). The resulting image file is available for download from the server for 3 days.

If desired, the PostScript© output file can be manually modified (e.g., to include additional features, labels) by importing it into PostScript© editor software (like Adobe Illustrator or the freeware tool Inkscape: http://www.inkscape.org/). However, for routine applications this will not be necessary.

Nomenclature conventions

According to the existing nomenclature conventions, the names of organelle-encoded genes are composed of three lowercase letters defining the protein or protein complex followed by an upper case letter or an Arabic numeral defining the subunit encoded by the specific gene. For
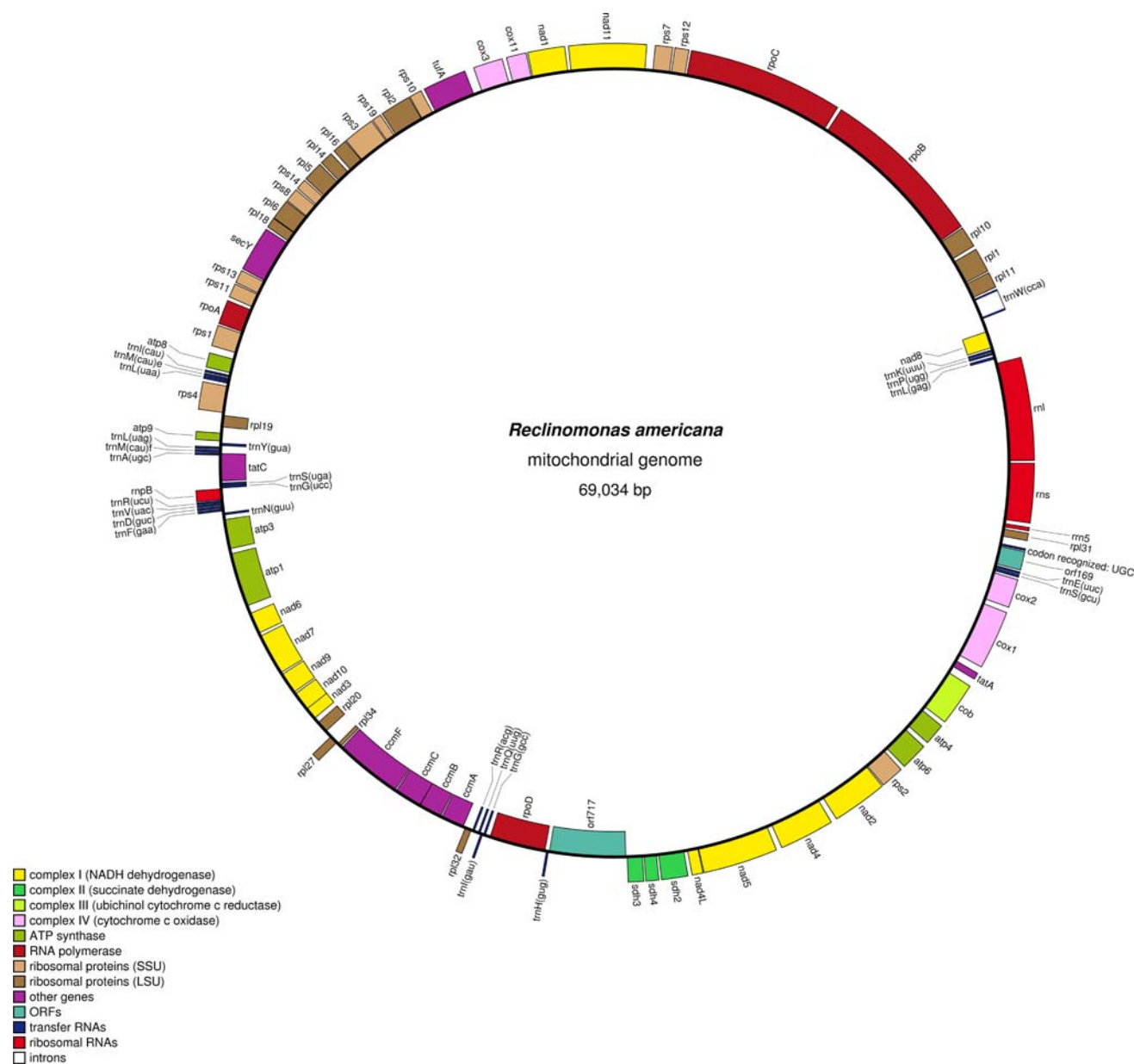
Fig. 3 Physical map of the mitochondrial genome of the protist *Reclinomonas americana* (AF007261.1) as drawn by OGDRAW. This is the most gene-rich mitochondrial genome known to date and considered to represent an ancestral type of mtDNA (Lang et al. 1997). Genes *inside of the circle* are transcribed clockwise, genes *outside the circle* are transcribed counterclockwise. See the physical map in the original publication for comparison (Lang et al. 1997). The map was drawn using the standard settings of OGDRAW, without activating or inactivating any features. Activation of the 'tidy-up' option would change the tRNA names to the proposed consensus nomenclature (e.g., change 'trnS(gcu)' to 'trnS-GCU'). Two minor annotation errors should be changed by editing the GenBank entry in a text editor: (1) The RNA subunit of RNase P (rnpB; at approximately nine o'clock) is misannotated as an rRNA in the GenBank entry (AF007261.1). (2) The odd gene name 'codon recognized: UGC' (at approximately three o'clock) is because the authors of the GenBank entry forgot to enter the gene name 'trnC(ugc)' for this tRNA gene (tRNA-Cys)

example, the plastid gene *psbA* is specified as a photosystem II subunit by the abbreviation *psb*, whereas *A* denotes the gene product as the reaction center protein D1. Similarly, *rps12* is composed of the three-letter abbreviation for the <u>r</u>ibosomal <u>p</u>roteins of the <u>s</u>mall subunit and the number of the subunit (S12). These nomenclature conventions are usually correctly applied to protein-coding genes in GenBank entries of plastid and mitochondrial genomes. However, they are unfortunately not generally obeyed in the case of RNA genes, such as transfer RNA (tRNA) and ribosomal RNA (rRNA) genes. Screening through GenBank entries, we have found a great variety of different gene name formats for such non-protein-coding genes. For example, the tRNA for tryptophan can be found annotated

```
<config>
      <feature>
            <color>151,190,13</color>
            <fullname>ATP synthase testclass</fullname>
            <drawflag>1</drawflag>
            <pattern>^atp.*</pattern>
            <type>gene</type>
      </feature>

      <feature>
            <color>0,102,44</color>
            <fullname>Photosystem I testclass</fullname>
            <drawflag>0</drawflag>
            <pattern>^psa.*</pattern>
            <type>CDS</type>
      </feature>
      [...]

</config>
```

Box 1 Example structure of a custom configuration file. For each feature class, a set of parameters describing the characteristics of the class can be defined in an XML-like format. The color in which a given feature class (e.g., gene class) shall be drawn in the map is specified between the <color>-tags as red, green and blue (RGB) values. The <drawflag> can take the values 1 or 0, defining whether the feature class shall be included in the map (1) or not (0). The <pattern> tag encloses a Perl regular expression that should match the names of all features that shall be included in the class. For example, in the first set of parameters, all "gene"-type features that start with the characters "atp" followed by any number of arbitrary characters will be categorized as "ATP synthase testclass" features. The <type> tag defines which type the features in a given class must have. For an exhaustive list of possible feature types in GenBank entries, see http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html. A template configuration file is available for download under: http://ogdraw.mpimp-golm.mpg.de/OGDrawconf_template_linux.xml (Linux version) or http://ogdraw.mpimp-golm.mpg.de/OGDrawconf_template_win.xml (Windows version)

as 'tRNA-Trp (UGG)' or 'tRNA-Trp anticodon: UGG' or 'tRNA-Trp codon recognized: UGG', just to name a few of the many different names used. To unify the naming of tRNA genes and provide the shortest possible yet unambiguous description, OGDRAW applies the following format: *trn(AMINOACID ONE-LETTER CODE)-(ANTICODON)*. Thus, the tRNA for alanine with the anticodon UGC is named *trnA-UGC*. Analogously, the denomination *rrn(SIZE IN SVEDBERG UNITS)* is used to specify rRNA genes (e.g., *rrn16* for the 16S ribosomal RNA). Many of the organellar genomes deposited recently in GenBank already follow these guidelines. In all other cases, the 'tidy-up' option can be activated, and OGDRAW will then try to reformat the names to match the above format. Alternatively, the GenBank entry can be edited manually to correct odd gene names (see subsequently, and Figs. 2, 3).

Use of the program

The OGDRAW webtool providing a simple-to-use interface and additional information for the user is available at the following URL: http://ogdraw.mpimp-golm.mpg.de/. The successful use of OGDRAW requires no special knowledge and is largely self-explanatory. The program offers extremely simple and user-friendly input options. All input information required is entered in just two simple screens: (1) a first screen to enter either a GenBank accession number or upload a file and (2) a second screen allowing the user to define the desired output options (Fig. 1). The program's output is highly customizable with respect to the features to be included in the map (e.g., gene classes, restriction enzyme cleavage sites, inverted repeat regions, GC content graph) as well as the output file type (Fig. 1). Useful standard features are activated by default so that an off-the-shelf map can be created directly by a single click on the 'Create map' button (Fig. 1). In addition to the PostScript© vector image format, OGDRAW offers an array of raster image formats (TIFF, JPEG, GIF and PNG) and also allows the user to choose the desired resolution. A 600 dpi high-resolution publishing quality map is usually computed in just a few seconds.

OGDRAW uses inner/outer circle depiction to visualize the transcriptional orientation of genes. Genes annotated in the GenBank entry as 'complement' are shown in the inner circle and transcribed clockwise (Figs. 2, 3). All genes transcribed counterclockwise are shown in the outer circle.

A limited number of mitochondrial DNAs map as linear genomes. OGDRAW recognizes linear genomes and automatically draws them as linear maps. An additional option available for both linear and circular genomes is to zoom into the genome and let OGDRAW display only a pre-defined region.

Although even first-time users will not need more than a few minutes to obtain a genome map with OGDRAW, it is

important to realize that the quality of the map content depends on the quality and correctness of the input file. If the provided GenBank file contains annotation errors, awkward features or odd gene names not conforming to the standard nomenclature for organellar genes (see above), these problems are likely to be carried over into the map displayed by OGDRAW. The frequently asked questions (FAQ) section of the OGDRAW web site (http://ogdraw.mpimp-golm.mpg.de/) provides solutions for problems caused by inaccuracies in GenBank files. Odd gene names, for example, can be easily dealt with by downloading the GenBank file, opening it in a text editor like 'Notepad' (Windows), 'TextEdit' (Mac) or 'Kate' or 'Gedit' (Linux), searching for the odd name and directly editing it in the copy of the GenBank file (Figs. 2, 3). If then the modified version of the file is uploaded to OGDRAW, the corrected map will display. Another problem that occasionally arises comes from unconventionally annotated split genes. For example, in the latest GenBank entry of the *Arabidopsis thaliana* mitochondrial genome (Y08501.2), intron-containing genes are not annotated as single contiguous 'gene' loci, but instead are annotated as split loci. (Normally, one would annotate the 'gene' locus contiguously from the start to the stop codon and annotate the split gene structure under the 'CDS', 'exon' and 'intron' features in the entry.) The annotation is correct in the genome version deposited in the RefSeq database (NC_001284; with the exception of the misannotated trans-spliced *nad1* locus), which therefore displays better than the latest GenBank version (cp. Unseld et al. 1997). If the latest GenBank version is to be used for drawing a physical map with OGDRAW, the 'gene' features of intron-containing genes must be appropriately edited in a text editor. This can be done very easily, because it is sufficient to delete the information on the exon–intron structure from the 'gene' line.

Finally, OGDRAW can also be used to detect gross errors in GenBank files prior to submission to the database. For example, in the tobacco plastid genome sequence (Z00044.2) in GenBank, the *psbB* operon is misannotated as *petD* gene. This error becomes obvious when running the GenBank file through OGDRAW, because the misannotated *petD* gene now overlaps with all genes of the *psbB* operon in the map. (Note that, for the tobacco ptDNA, the 'tidy-up' option should be activated to correctly display the trans-spliced *rps12* locus.)

To offer maximum flexibility, the core of the program has been designed as a set of Perl modules (GeneMap) providing an object-oriented interface to all functions and offering a slightly greater degree of customizability than the Web interface. This design enables the user to incorporate the module in custom-made Perl scripts. The modules, including full documentation, can be obtained from the authors upon request.

Conclusions

OGDRAW provides the community with a convenient tool to generate high-quality genome maps of organellar genomes. To ensure general availability, the program has been set up on a publicly accessible web server. Genbank entries can be provided via an accession number or can be uploaded directly as GenBank flat files enabling the user to quickly generate maps of multiple organellar genomes for direct visual comparison. Moreover, the possibility to upload a sequence file offers the opportunity to modify or update the original GenBank entry or to draw maps from to-date unpublished genomics data.

OGDRAW has been specifically developed and optimized for the display of organellar genomes (ptDNAs and mtDNAs) that map as circular molecules. Special care has been taken to include and correctly display the structural and genetic peculiarities of these genomes, for example, the presence of large inverted repeat regions and the gene organization in polycistronic transcription units (operons). The program's output is highly customizable concerning the features to be included in the map and offers both vectorized (PostScript®) and rasterized image formats at a wide range of resolutions, including, for example, 600 dpi publication-quality TIFF. The frequently asked questions (FAQ) section of the web site offers solutions to problems that may arise. The webtool providing an easy-to-use interface, and supplemental information is available at the following URL: http://ogdraw.mpimp-golm.mpg.de/.

Future developments of the OGDRAW program could include the design of interactive (clickable) maps to display all available information for individual genes, the possibility to link genes to their database entries (including, for example, protein sequences and structures) and display options for RNA editing sites and intron types.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–464

Backert S, Nielsen BL, Börner T (1997) The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. Trends Plant Sci 2:477–483

Bock R (2007) Structure, function, and inheritance of plastid genomes. Top Curr Genet 20:29–63

Bungard RA (2004) Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. Bioessays 26:235–247

Burland TG (2000) DNASTAR's Lasergene sequence analysis software. Methods Mol Biol 132:71–91

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of Pelargonium × hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175–2190

dePamphilis CW, Palmer JD (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348:337–339

Dong X, Stothard P, Forsythe IJ, Wishart DS (2004) PlasMapper: a web server for drawing and auto-annotating plasmid maps. Nucleic Acids Res 32:W660–W664

Douglas SE, Penny SL (1999) The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. J Mol Evol 48:236–244

Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Tumel M, Brossard N, Delage E, Littlejohn TG, Plante I, Rioux P, Saint-Louis D, Zhu Y, Burger G (1998) Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res 26:865–878

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet 46:123–139

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387:493–497

Lu G, Moriyama EN (2004) Vector NTI, a balanced all-in-one sequence analysis suite. Brief Bioinform 5:378–388

Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, Stern DB (2002) The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. Plant Cell 14:2659–2679

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S-i, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. Nature 322:572–574

Palmer JD, Nugent JM, Hebron LA (1987) Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proc Natl Acad Sci USA 84:769–773

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Stothard P, Wishart DS (2005) Circular genome visualization and exploration using CGView. Bioinformatics 21:537–539

Tsudzuki T (2000) A graphic tool for circular genome maps. Nucleic Acids Symp Ser 44:189–190

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet 15:57–61

Wakasugi T, Tsudzuki T, Sugiura M (2001) The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. Photosynth Res 70:107–118