MORPHEUS, a Webtool for Transcription Factor Binding Analysis Using Position Weight Matrices with Dependency.

Eugenio Gómez Minguet, Stéphane Segard, Céline Charavay, François Parcy

PLOS | ONE

RESEARCH ARTICLE

MORPHEUS, a Webtool for Transcription Factor Binding Analysis Using Position Weight Matrices with Dependency

Eugenio Gómez Minguet[1¤], Stéphane Segard[2], Céline Charavay[2], François Parcy[1]*

1 Laboratoire de Physiologie Cellulaire Végétale, Unité Mixte de Recherche 5168, Centre National de la Recherche Scientifique, Commissariat à l'Énergie Atomique, Institut National de la Recherche Agronomique, Université Joseph Fourier Grenoble I, 38054, Grenoble, France, **2** Laboratoire de Biologie à Grande Echelle, CEA/INSERM U1038/UJF—Grenoble 1, IRTSV, F-38054, Grenoble, Cedex 9, France

¤ Current address: Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Universidad Politécnica de Valencia, Avda de los Naranjos s/n, Valencia, 46022, Spain
* francois.parcy@cea.fr

Citation: Minguet EG, Segard S, Charavay C, Parcy F (2015) MORPHEUS, a Webtool for Transcription Factor Binding Analysis Using Position Weight Matrices with Dependency. PLoS ONE 10(8): e0135586. doi:10.1371/journal.pone.0135586

Editor: Frances M Sladek, Univeristy of California Riverside, UNITED STATES

Received: April 19, 2015

Accepted: July 24, 2015

Published: August 18, 2015

Data Availability Statement: All relevant data are within the paper, its Supporting Information files and in Morpheus webpage: http://biodev.cea.fr/morpheus/.

Funding: This work was supported by the Charmful Program (ANR11-BSV2-005-01) from the Agence Nationale de la Recherche Program to FP and an ATIP+ Program from the Centre National de la Recherche Scientifique to FP and EGM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Transcriptional networks are central to any biological process and changes affecting transcription factors or their binding sites in the genome are a key factor driving evolution. As more organisms are being sequenced, tools are needed to easily predict transcription factor binding sites (TFBS) presence and affinity from mere inspection of genomic sequences. Although many TFBS discovery algorithms exist, tools for using the DNA binding models they generate are relatively scarce and their use is limited among the biologist community by the lack of flexible and user-friendly tools. We have developed a suite of web tools (called Morpheus) based on the proven Position Weight Matrices (PWM) formalism that can be used without any programing skills and incorporates some unique features such as the presence of dependencies between nucleotides positions or the possibility to compute the predicted occupancy of a large regulatory region using a biophysical model. To illustrate the possibilities and simplicity of Morpheus tools in functional and evolutionary analysis, we have analysed the regulatory link between LEAFY, a key plant transcription factor involved in flower development, and its direct target gene *APETALA1* during the divergence of Brassicales clade.

Introduction

The binding of transcription factors (TF) to *cis* elements is a key component of most biological processes. Being able to detect TF binding sites (TFBS) by inspecting genome sequences helps understanding how organisms work and how they evolved. Methods based on Chromatin Immunoprecipitation (ChIP) such as ChIP-Chip [1], ChIP-Seq [2] or ChIP-exo [3] allow the identification of all genomic regions bound by a given TF in one experimental condition and suites as Bedtools [4, 5] offer many tools to manipulate them. To precisely identify the TFBS present in these regions, estimate their affinity, predict binding sites that might be bound in

Competing Interests: The authors have declared that no competing interests exist.

other experimental conditions, or study organism where ChIP experiments are more challenging, TF DNA binding models are extremely useful. There are multiple ways to model TFBS. The most common is the Position Weight Matrix (PWM) that, for each sequence, computes a score directly related to the TF/DNA affinity ([6] for a review). This method, however, assumes that each base of the BS contributes independently to the affinity of the TF for DNA [7] and there is evidence that interdependencies between positions exist [8, 9] and that taking into account dinucleotide dependencies between two adjacent positions already improves predictions [10]. Several alternatives with specific advantages have been proposed using nucleotide subsequences (K-mers) rather than mononucleotide positions [8, 11, 12] or hidden Markov models (HMM) [13]. However, in most cases, PWMs provide simple and reliable estimation of binding affinity [14]. We propose to adapt the convenient PWM model by adding dependency information at specific positions of the matrix. As documented for several TFs [10, 15–17], this will improve the prediction power of some PWM models.

Although several tools such as RSAT [18], PROMO [19], MatInspector [20] or LASAGNA [21] are available to identify overrepresented motifs in a set of sequence and build binding models, none of them allows using PWM with dependencies nor to calculate occupancy of DNA regions using biophysical models [22]. We have developed a new algorithm that uses PWM with any combination of dependent and independent positions. We incorporated it in a user friendly set of tools called MORPHEUS, which offers several specific advantages over existing tools: 1) it is a web tool that does not require any programming skill and can thus be widely used by the biologist community, 2) users can import their own matrices, not only those found in databases, 3) position interdependencies can be included between any positions of the matrix and in combination with independent positions, a possibility currently offered by none of the existing web tools, 4) a global "predicted occupancy" value can be computed for whole DNA regions using a biophysical model [22] that integrates the presence of individual binding sites.

Results and Discussion

Morpheus Matrix Format and *m*PWM algorithm

The Morpheus PWM format (*m*PWM) allows the introduction of information on di- or trinucleotide dependencies between any indicated positions (not just adjacent ones) within a binding site. Unlike other models that increase model complexity for all positions, *m*PWM conserves the simplicity of a PWM except for interdependent positions. Using *m*PWM, interdependencies are defined as additional $4^{(d)}$ matrices (d = 2 for dinucleotide dependency, d = 3 for triplets) for any position combination (Example matrix files are provided as S1 Text and S2 Text).

mPWM Format Conversion Tool

We have provided a tool to generate *m*PWM from an alignment of transcription factor binding sites. Positions with dependency have to be detected using programs such ENOLOGOS [23] and provided as a list of dependent positions for the conversion tool to automatically generate the corresponding $4^{(d)}$ matrices. Depending on the TF structural features, the possibility is offered to generate symmetric matrices.

Morpheus tools

The Morpheus suite allows the calculation of relative affinity of TFBS from *m*PWM. Based on the scores of individual binding sites present in a large DNA region, Morpheus also computes

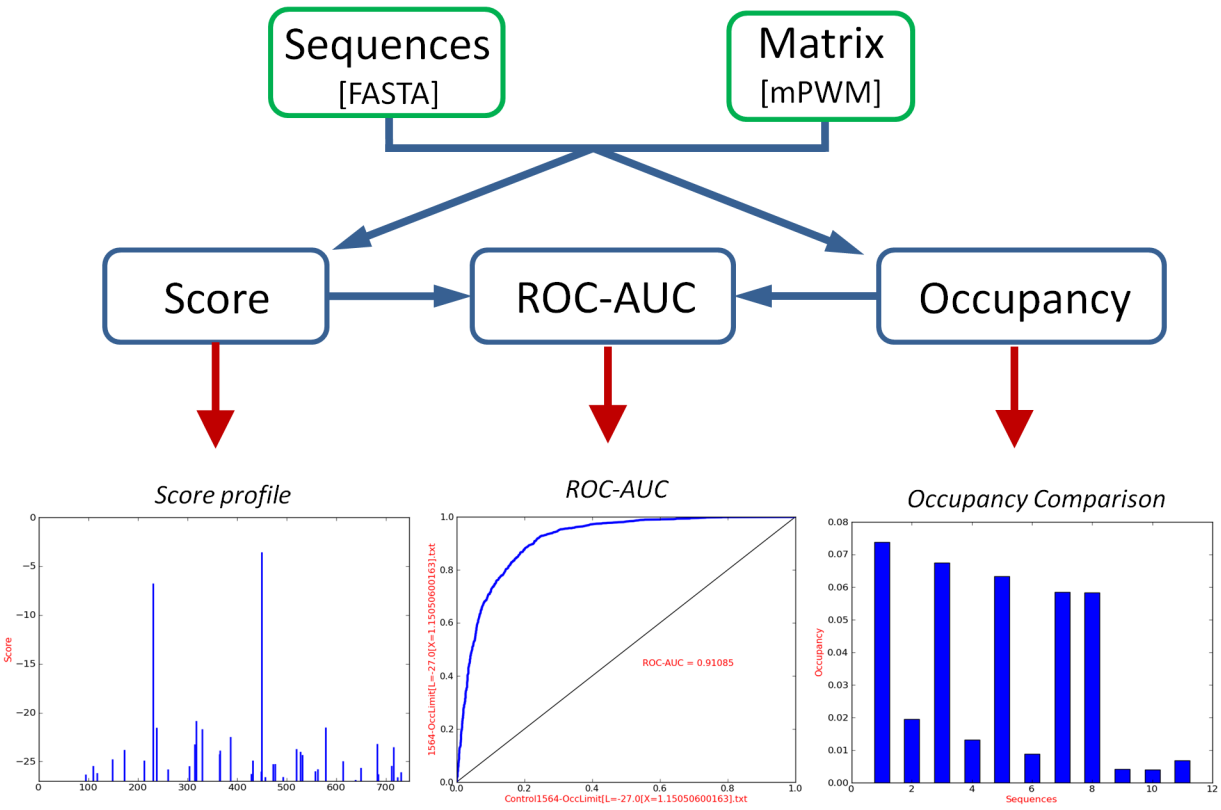

Fig 1. Morpheus flowchart and example of result representation. The tool Score scans DNA regions and computes the scores of TFBS. The bottom left graph shows TFBS locations and scores; such score profile is generated for each sequence submitted. The Occupancy tool computes the TF predicted occupancy of each DNA region taking as input sequence files (in fasta format) and a binding model information (*m*PWM format). Complete results are written in text files and also displayed as graphical outputs for quick results overview. Bottom right panel shows a occupancy comparison between different DNA regions. The bottom central panel illustrates the ROC-AUC curve and value obtained with the ROC-AUC tool.

doi:10.1371/journal.pone.0135586.g001

the predicted occupancy using a biophysical model as previously described [22, 24, 25]. This possibility, not offered by any other web-tool, is particularly important as individual *cis* elements can vary within a regulatory region even though the occupancy and overall regulation are conserved [16, 26]. The predicted occupancy thus offers a global measure that allows comparing regions independently of the individual binding site variations.

Morpheus webtool it is composed of three tools:

- Morpheus 'Score' tool scans DNA regions and computes the scores of individual TFBS. The user can choose to display only the TFBS of highest score of each region, the TFBS with a score higher than a given threshold score or all TFBS. For an easy graphical representation, this tool also generates score profiles for each sequence as well as an histogram with all scores (Fig 1).

- Morpheus 'Occupancy' tool computes the TF predicted occupancy of each DNA region using formalism described above [22, 24] with the option of using only the scores exceeding a given threshold. Occupancy calculation is based on the correlation between predicted score and relative dissociation constant which can be obtained from *in vitro* measurements of relative affinities [27]. If this data is not available a relative occupancy can be calculated using default values for the parameters.

Both score and occupancy options take two files as input: a file with sequences in fasta format and a *m*PWM.

- Morpheus 'ROC' allows assessing the quality of a TF binding model by performing a Receiver Operating Characteristics (ROC) analysis [28]. This analysis measures the discriminative power of a TF matrix by comparing a set of bound regions (obtained for example from ChIP-Seq experiments) to a negative control set generated by the user. The comparison uses either the best score TFBS of each sequence or its occupancy and the Area Under the Curve (AUC) value is computed as a measure of the model predictive power.

All three programs display graphic output (Fig 1) for quick results overview and text files with complete results for further analysis by the user.

For illustration of how Morpheus suite works, we present here a set of analyses performed with the LEAFY (LFY) protein, a plant TF with a central role in the evolution and development of flowers [29, 30]. According to *in vitro* affinity measurement a PWM has been proposed for this factor (LFY-Trip) that includes three dependency triplets in a symmetric motif of 19 positions [16], in accordance with the information obtained from the LFY-DNA crystal structures [31, 32].

LEAFY Binding model evaluation using ROC

The availability of ChIP-Seq data allows performing ROC analysis using the described set of bound genomic regions [16] as well as a negative set of non bound regions (see Methods for description of negative set generation) to compare the predictive power of this matrix against the previously described consensus motifs [31, 33, 34]. To do this, Scores or Occupancies are computed for both the positive and negative sets with each binding model (using Score or Occupancy tools) and the result serves as input for the ROC program. A histogram is generated that represents the distribution of scores or occupancy values for each data set (Fig 2A). This tool also generates an image file with the ROC curve and a text file with all the data. In Fig 2B, we use ROC-AUC results to illustrate the increased prediction power of the LFY-Trip PWM as compared to previously used consensus sequences. This tool can be used to compare the various matrices identified by various motifs finding algorithms in order to select the one with the best predictive power. Next, we illustrate how the LFY *m*PWM can be used for functional or evolutionary analysis of a regulatory relationship using the Score and Occupancy tools.

Prediction of LEAFY binding sites on *APETALA1* promoter

We focused on the link between *LFY* and its direct target *APETALA1* (*AP1*) involved in the development of flowers [29, 30]. The MADS box TF gene *AP1* arose from duplication of the *FRUITFUL* (*FUL*) gene and this event was proposed to be important in the fixation of flower structure in eudicot plants [35]. *AP1* have also experienced a more recent Brassicaceae-specific duplication [36, 37] generating the *CAULIFLOWER* gene. While *AP1* is a direct target of LFY, there is no evidence for direct regulation of *FUL* or *CAL* [16, 38]. We illustrate here how Morpheus can be used to explore *LFY-AP1* link through eudicot plants evolution.

The functional analysis of *AP1* promoter and its regulation by LFY binding has been performed in the model plant *Arabidopsis thaliana*. A few promoter versions have been tested *in vivo* [39] including different promoter lengths (2.2, 1.7, 0.9 and 0.6 kb), and mutations in three candidate LFYBS (bs1, bs2 and bs3) displaying consensus motifs [31, 33, 34]. The score profile of *AP1* promoter generated with Morpheus Score tool (option "limit = -25") illustrates the position of the best LFYBS in *AP1* promoter (Fig 3A). We computed Occupancy values for all promoter versions and compared these values to the *in vivo* activity of the corresponding promoter fragment (Fig 3B). *In vivo*, mutations in bs2 and bs3 had weak effect while mutation in bs1 had a strongest effect, which is in accordance with their computed scores and not with the presence of the consensus sequence. The good correlation between the two types of data

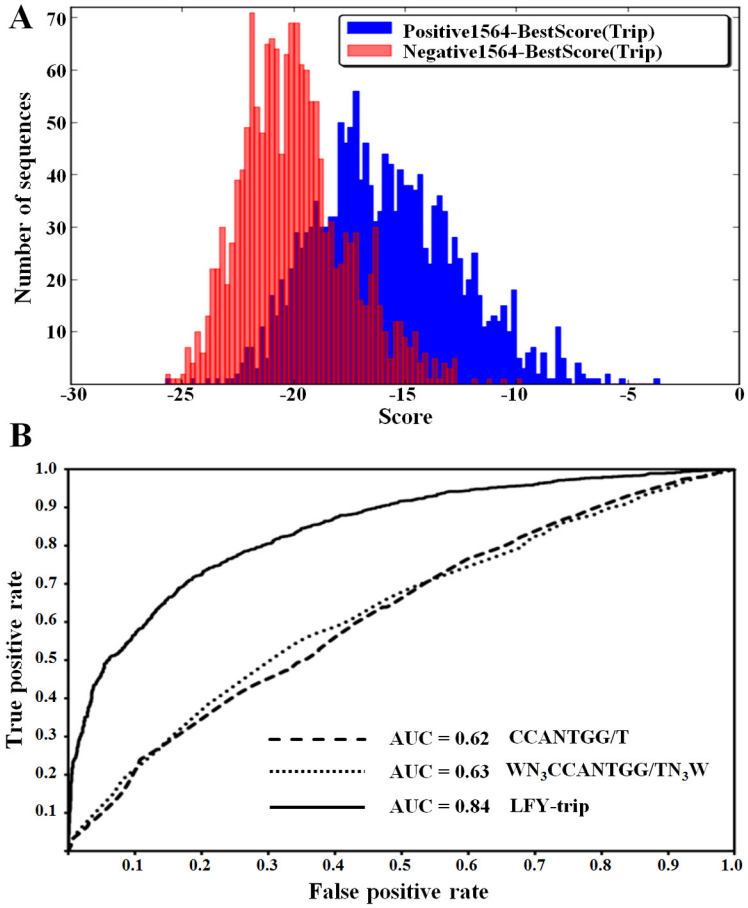

Fig 2. The performance of a TF model can be evaluated by its capability to discriminate between bound and non-bound regions as determined from a ChIP-Seq experiment. The Morpheus 'ROC' tools computes the ROC-AUC value as a measure of the model predictive power. **A)** The histogram graphical output displays the distribution of score values for the best binding sites present on each DNA sequence. **B)** ROC data output for three binding models: two consensus motifs and LFY-trip (input data has been generated using the Morpheus 'Occupancy' tool). The LFY-trip model largely outperforms the two consensus models.

doi:10.1371/journal.pone.0135586.g002

illustrates the power of the biophysical model to predict the impact of TFBS changes on gene expression by integrating all possible TFBS present in a regulatory region.

Transcriptional regulation and evolution

Next, we use the Morpheus tools to study the evolution of the link between LFY and genes of the *FUL* clade (*AP1*, *CAL* and *FUL*). Scanning of 2 kb of promoter sequences in various species illustrates well the diversity of LFYBS landscapes (Fig 4B). As it is difficult to draw clear conclusions directly from these TFBS profiles, we computed the Occupancy (Occ) for these different promoters (Fig 4A). We found a higher occupancy of the *AP1* ortholog promoters as compared to those of *CAL* and *FUL* in the Brassicales clade, a result in good accordance with experimental data available in Arabidopsis [16, 33, 38, 40]. This analysis suggests that the link between LFY and *AP1* originated before the divergence of *B. rapa*.

Interestingly, the promoter of the *AP1* gene from the Brassicale *Carica papaya* displays a low occupancy though evidence suggests a regulation by LFY. We wondered whether this

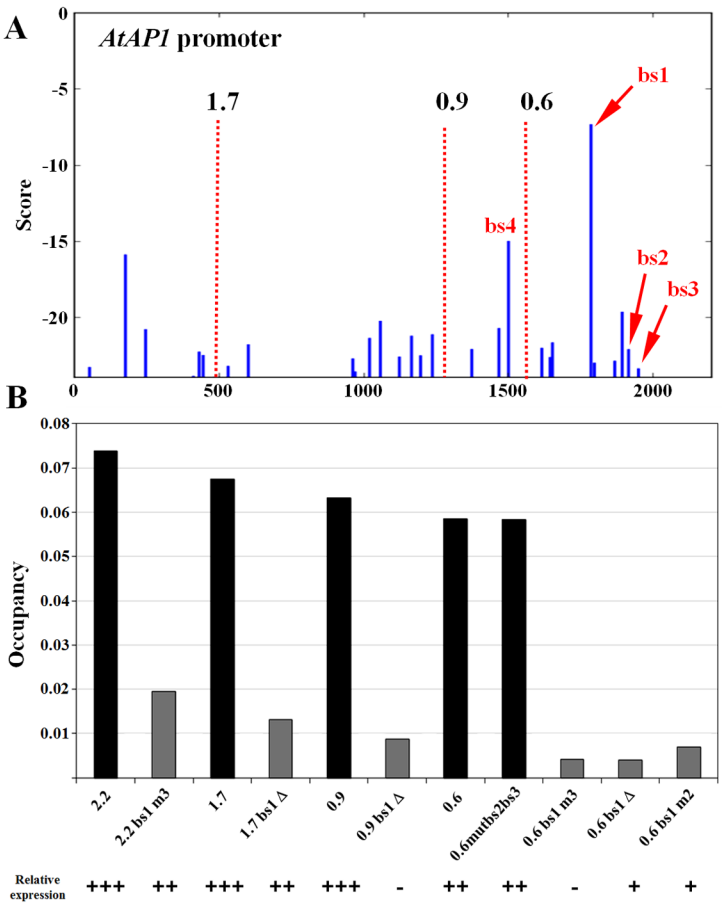

Fig 3. Comparison of LEAFY binding analysis in *A. thaliana AP1* promoter using the Morpheus suite with *in vivo* promoter expression study [39]. A) Score profile graphic output of the Morpheus 'Score' tool (option limit = -25) using 2.2 kb upstream of *AP1* start codon. Red dotted lines show the different promoter sizes and arrows mark the mutated BS, accordingly with promoters set described in [39]. **B)** Predicted occupancy (option All) shows a good correspondence with relative expression of each promoter version as determined experimentally in a published study and summarized: expression levels: +++ (high), ++ (medium), + (low),—(not detectable). The number indicates the size of the promoter (2.2, 1.7, 0.9 or 0.6 kb), m2 and m3 indicate mutations in bs2 and bs3 respectively, Δ1 indicates a deletion of bs1. *In vivo*, mutations of bs2 and bs3 (promoter 0.6 mutbs2bs3) has only a weak effect while elimination of bs1 drastically affects *AP1* expression.

doi:10.1371/journal.pone.0135586.g003

reflected an absence of regulatory link between LFY and *AP1* in this species or whether the LFY binding sites could be located outside of the promoter. We thus scanned the region downstream of the start codon and we found a predicted BS with a very good score that could be responsible for the *AP1* regulation by LFY in this species despite the absence of high affinity LFYBS in the promoter. None of the other species with low promoter occupancy displayed this 3' binding site (data not shown). This data supports the hypothesis that LFY-*AP1* link originated before Brassicales divergence. The low occupancy values for *AP1* promoters in Fabids suggest LFY does not regulate *AP1* in these species. However, because there are intermediate Occupancy values in the Fabaceae clade, a more detailed experimental characterization would be required in these species to assay the possible existence of a regulatory relationship between LFY and *AP1*.

These analyses illustrate how genomic sequences can be analysed with the Morpheus tool to generate hypotheses regarding gene regulation and regulatory network evolution. More

examples can be found in three additional studies [16, 41, 42] that used Morpheus while under development. As more TF binding models become available, such tools will become increasingly important to exploit the genomic data, answer evolutionary questions and bringing up new working hypotheses.

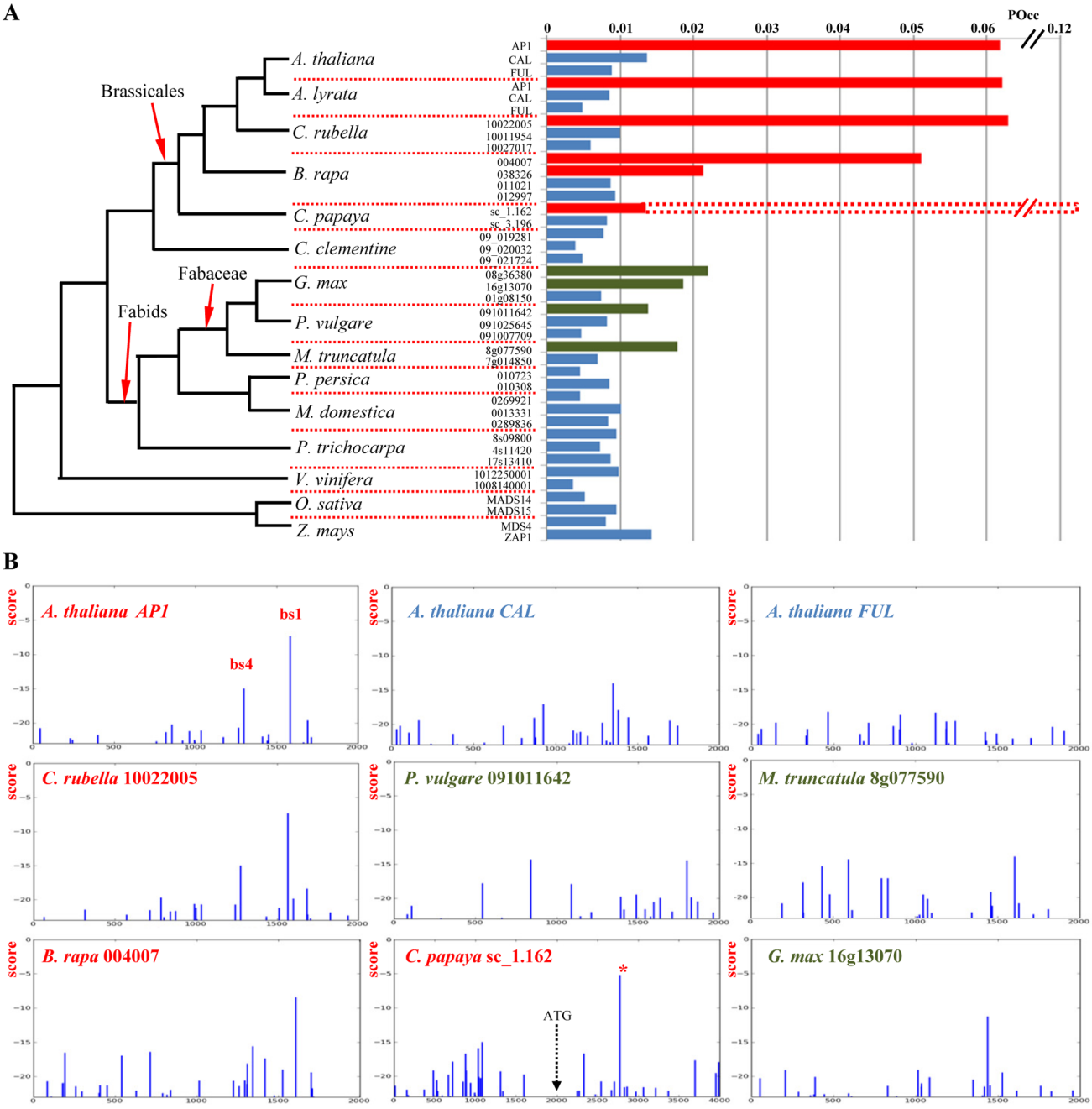

Fig 4. Evolutionary analysis of LFY binding on *AP1* promoters. Genomic sequences were obtained from the Phytozome database and 2 kb promoter upstream the ATG were used. Only well annotated genes were used. **A)** Predicted occupancy (Option limit = -23) for each promoter. Phylogenetic relationships between species are represented. **B)** Score profile (limit = -23) of some representative promoters. The higher occupancy for *AP1* promoters in Brassicales (red) suggests that the regulatory link between LFY and *AP1* in *A. thaliana* arose before the divergence of this clade. Interestingly, *C. papaya* with low occupancy in *AP1* promoter has a candidate BS of very good score downstream the start codon likely to be responsible for a regulation by LFY. In Fabids, the low occupancy values suggest that LFY does not regulate *AP1*, though some promoters have intermediate occupancy values (green) what will need further analysis.

doi:10.1371/journal.pone.0135586.g004

Conclusions

Morpheus web allows a user-friendly suite of tools for the calculation of TFBS relative affinity on DNA sequences. It incorporates unique features such as dependency between specific positions, occupancy calculation and ROC-AUC estimation that do not exist in any currently available webtool. We have illustrated how it can be used to infer hypothesis about TFBS functional significance or about evolution of regulatory links. Experienced users can download Morpheus scripts code for specific purpose, however no programming skills are needed to use Morpheus web-tools. With all its unique characteristics and with the possibility of using any own-modified *m*PWM, we believe Morpheus should have strong acceptance among biologists. Morpheus web-tools, complete user guide and downloading versions are available at Morpheus website: http://biodev.cea.fr/morpheus/.

Methods

Morpheus

All scripts for Morpheus tools are written in Python programming language (ver 2.6.7). The graphic output requires two modules: Numpy (http://numpy.scipy.org/) and Matplotlib (http://matplotlib.sourceforge.net/). Morpheus tools are available in the Morpheus web (http://biodev.cea.fr/morpheus/), as well as downloading versions with or without graphic output, user guide and complete descriptions. The web is hosted and maintained by the GIPSI team (CEA Saclay).

When the score matrix is not directly provided, Morpheus computes it based on 'Count' or 'Frequency' matrices using $W_{n,i} = Ln(f_{n,i}/f_{max,i})$ where $W_{n,i}$ is the weight at position i for nucleotide n, $f_{n,i}$ is the frequency of nucleotide n at position i and $f_{max,i}$ is the maximal frequency observed at position i [43]. Each $4^{(d)}$ dependency matrix is preceded by a line indicating the positions involved (S2 Text). For score calculation the *m*PWM algorithm first get the value for each independent position from the independent matrix and then for all the dependency combinations from the $4^{(d)}$ matrices. If *in vitro* affinity data is available to correlate score with relative dissociation constant, the correlation values can be indicated in *m*PWM file or, if they are not indicated, the program will use default parameters (corresponding to a line curve with scope equal to one). From matrix file, *m*PWM algorithm first identifies the list of independent positions (i) and the list of dependent positions groups (j; each one associated with a $4^{(d)}$ matrix), then the score of each DNA sequence is calculated as:

$$Sequence_{score} = \sum_{p \in i} score_{p_{nt}} + \sum_{q \in j} score_{q_{dep}}$$

where $score_{p_{nt}}$ is the score in the position p for the nucleotide nt (A,C,G or T) in the independent matrix and $score_{q_{dep}}$ is the score in the $4^{(d)}$ matrix of group q for the sequence combination dep (dinucleotide or triplet). Example of matrix files in Morpheus format with or without dependencies are provided as S1 Text and S2 Text, respectively.

Occupancy calculation (default parameters)

Occupancy calculation is based on the relation between predicted score and relative dissociation constant [16, 24], score = -a * ln(Kd) + b. A and b values can be provided in the *m*PWM file when they have been determined. If not, Morpheus will use default parameters (a = 1.0 and b = 0.0). Occupancy calculation formalism also requires the TF concentration [X], which can optionally be indicated if available. Since this value is rarely available, as default, Morpheus

uses $[X] = e^{(b/a)}$ corresponding to a TF concentration at which the best possible binding site (maximal score) is bound with a probability of 0.5.

Sequences sets for ROC-AUC calculation

Positive bound sequences was taken from [16] (S3 Text). To generate a negative set of sequences for ROC-AUC analysis, we have randomly selected in the *A*. *thaliana* genome a set of sequences that do not overlap with the positive set and with the same size distribution (S4 Text).

FUL clade genomic sequences

Genomic sequences were obtained from Phytozome database [44] by Blast search using the protein sequence of *AtAP1* (At1g69120), *AtCAL* (At1g26310) and *AtFUL* (At5g60910). Hits without transcripts information or with incomplete gene prediction were discarded. A region of 2 kb upstream the ATG were used for relative binding score calculation. All sequences used in this study can be found in S5 Text.

Supporting Information

S1 Text. Example of mPWM format without dependency.
(TXT)

S2 Text. Example of mPWM format with dependency.
(TXT)

S3 Text. Positive Sequences Set.
(TXT)

S4 Text. Negative Sequences Set.
(TXT)

S5 Text. FUL clade genomic sequences.
(TXT)

Acknowledgments

The authors thank Miguel A. Blázquez, David Alabadí, Paco Madueño and Cristina Ferrándiz for discussion, the GIPSI team (CEA Saclay) especially A. Martel for technical support and for hosting web site and A. Mathelier for its critical reading of the manuscript. This work was supported by the Charmful program (ANR11-BSV2-005-01) from the Agence Nationale de la Recherche program to FP and an ATIP+ program from the Centre National de la Recherche Scientifique to FP and EGM.

Author Contributions

Conceived and designed the experiments: EGM FP. Performed the experiments: EGM. Analyzed the data: EGM FP. Wrote the paper: EGM FP. Developed Morpheus webpage: SS CC. Developed Morpheus software: EGM.

References

1. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, et al. Genome-wide location and function of DNA binding proteins. Science. 2000; 290(5500):2306–9. Epub 2000/12/23. doi: 10.1126/science.290.5500.2306 PMID: 11125145.

2. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007; 316(5830):1497–502. Epub 2007/06/02. doi: 10.1126/science.1141319 PMID: 17540862.
3. Rhee HS, Pugh BF. ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Current protocols in molecular biology / edited by Frederick M Ausubel [et al]. 2012;Chapter 21:Unit 21 4. doi: 10.1002/0471142727.mb2124s100 PMID: 23026909; PubMed Central PMCID: PMC3813302.
4. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics. 2014; 47:11 2 1–2 34. Epub 2014/09/10. doi: 10.1002/0471250953.bi1112s47 PMID: 25199790; PubMed Central PMCID: PMCPMC4213956.
5. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26(6):841–2. doi: 10.1093/bioinformatics/btq033 PMID: 20110278; PubMed Central PMCID: PMC2832824.
6. Stormo GD. Modeling the specificity of protein-DNA interactions. Quantitative Biology. 2013; 1(2):115–30. doi: 10.1007/s40484-013-0012-4 PMID: 25045190
7. Stormo GD. DNA binding sites: representation and discovery. Bioinformatics. 2000; 16(1):16–23. PMID: 10812473.
8. Bulyk ML, Johnson PLF, Church GM. Nucleotides of transcription factor binding sites exert interdependent effects on the binding affinities of transcription factors. Nucleic acids research. 2002; 30(5):1255–61. doi: 10.1093/nar/30.5.1255 PMID: 11861919
9. Siddharthan R. Dinucleotide weight matrices for predicting transcription factor binding sites: generalizing the position weight matrix. PloS one. 2010; 5(3):e9722. doi: 10.1371/journal.pone.0009722 PMID: 20339533; PubMed Central PMCID: PMC2842295.
10. Zhao Y, Ruan S, Pandey M, Stormo GD. Improved models for transcription factor binding site identification using nonindependent interactions. Genetics. 2012; 191(3):781–90. doi: 10.1534/genetics.112.138685 PMID: 22505627; PubMed Central PMCID: PMC3389974.
11. Berger MF, Bulyk ML. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nature protocols. 2009; 4(3):393–411. doi: 10.1038/nprot.2008.195 PMID: 19265799; PubMed Central PMCID: PMC2908410.
12. Mathelier A, Wasserman WW. The Next Generation of Transcription Factor Binding Site Prediction. PLoS Comput Biol. 2013; 9(9):e1003214. doi: 10.1371/journal.pcbi.1003214 PMID: 24039567
13. Annala M, Laurila K, Lahdesmaki H, Nykter M. A linear model for transcription factor binding affinity prediction in protein binding microarrays. PloS one. 2011; 6(5):e20059. doi: 10.1371/journal.pone.0020059 PMID: 21637853; PubMed Central PMCID: PMC3102690.
14. Benos PV, Bulyk ML, Stormo GD. Additivity in protein–DNA interactions: how good an approximation is it? Nucleic acids research. 2002; 30(20):4442–51. doi: 10.1093/nar/gkf578 PMID: 12384591
15. Man T-K, Stormo GD. Non-independence of Mnt repressor–operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. Nucleic acids research. 2001; 29(12):2471–8. PMID: PMC55749.
16. Moyroud E, Minguet EG, Ott F, Yant L, Pose D, Monniaux M, et al. Prediction of regulatory interactions from genome sequences using a biophysical model for the Arabidopsis LEAFY transcription factor. The Plant cell. 2011; 23(4):1293–306. doi: 10.1105/tpc.111.083329 PMID: 21515819; PubMed Central PMCID: PMC3101549.
17. O'Flanagan RA, Paillard G, Lavery R, Sengupta AM. Non-additivity in protein-DNA binding. Bioinformatics. 2005; 21(10):2254–63. doi: 10.1093/bioinformatics/bti361 PMID: 15746285.
18. Medina-Rivera A, Defrance M, Sand O, Herrmann C, Castro-Mondragon JA, Delerce J, et al. RSAT 2015: Regulatory Sequence Analysis Tools. Nucleic acids research. 2015. doi: 10.1093/nar/gkv362 PMID: 25904632.
19. Farré D, Roset R, Huerta M, Adsuara JE, Roselló L, Albà MM, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic acids research. 2003; 31(13):3651–3. doi: 10.1093/nar/gkg605 PMID: 12824386
20. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, et al. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 2005; 21(13):2933–42. doi: 10.1093/bioinformatics/bti473 PMID: 15860560.
21. Lee C, Huang C-H. LASAGNA-Search: an integrated web tool for transcription factor binding site search and visualization. BioTechniques. 2013; 54(3). doi: 10.2144/000113999
22. Granek JA, Clarke ND. Explicit equilibrium modeling of transcription-factor binding and gene regulation. Genome biology. 2005; 6(10):R87. doi: 10.1186/gb-2005-6-10-r87 PMID: 16207358; PubMed Central PMCID: PMC1257470.

23. Workman CT, Yin Y, Corcoran DL, Ideker T, Stormo GD, Benos PV. enoLOGOS: a versatile web tool for energy normalized sequence logos. Nucleic acids research. 2005; 33(Web Server issue):W389–92. doi: 10.1093/nar/gki439 PMID: 15980495; PubMed Central PMCID: PMC1160200.

24. Roider HG, Kanhere A, Manke T, Vingron M. Predicting transcription factor affinities to DNA from a biophysical model. Bioinformatics. 2007; 23(2):134–41. doi: 10.1093/bioinformatics/btl565 PMID: 17098775.

25. Ward LD, Bussemaker HJ. Predicting functional transcription factor binding through alignment-free and affinity-based analysis of orthologous promoter sequences. Bioinformatics. 2008; 24(13):i165–71. doi: 10.1093/bioinformatics/btn154 PMID: 18586710; PubMed Central PMCID: PMC2718632.

26. Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science. 2010; 328(5981):1036–40. doi: 10.1126/science.1186176 PMID: 20378774; PubMed Central PMCID: PMC3008766.

27. Man TK, Stormo GD. Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. Nucleic acids research. 2001; 29 (12):2471–8. PMID: 11410653; PubMed Central PMCID: PMC55749.

28. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143(1):29–36. PMID: 7063747.

29. Liu C, Thong Z, Yu H. Coming into bloom: the specification of floral meristems. Development. 2009; 136(20):3379–91. doi: 10.1242/dev.033076 PMID: 19783733.

30. Moyroud E, Kusters E, Monniaux M, Koes R, Parcy F. LEAFY blossoms. Trends in plant science. 2010; 15(6):346–52. doi: http://dx.doi.org/10.1016/j.tplants.2010.03.007 PMID: 20413341

31. Hamès C, Ptchelkine D, Grimm C, Thevenon E, Moyroud E, Gérard F, et al. Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins2008 2008-10-08 00:00:00. 2628–37 p. doi: 10.1038/emboj.2008.184 PMID: 18784751

32. Sayou C, Monniaux M, Nanao MH, Moyroud E, Brockington SF, Thévenon E, et al. A Promiscuous Intermediate Underlies the Evolution of LEAFY DNA Binding Specificity. Science. 2014; 343 (6171):645–8. doi: 10.1126/science.1248229 PMID: 24436181

33. Parcy F, Nilsson O, Busch MA, Lee I, Weigel D. A genetic framework for floral patterning. Nature. 1998; 395(6702):561–6. doi: 10.1038/26903 PMID: 9783581.

34. Busch MA, Bomblies K, Weigel D. Activation of a floral homeotic gene in Arabidopsis. Science. 1999; 285(5427):585–7. PMID: 10417388.

35. Litt A, Irish VF. Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. Genetics. 2003; 165(2):821–33. PMID: 14573491; PubMed Central PMCID: PMC1462802.

36. Purugganan MD. The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. Journal of molecular evolution. 1997; 45(4):392–6. PMID: 9321418.

37. Purugganan MD, Suddith JI. Molecular population genetics of the Arabidopsis CAULIFLOWER regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95(14):8130–4. PMID: 9653152; PubMed Central PMCID: PMC20941.

38. Winter CM, Austin RS, Blanvillain-Baufume S, Reback MA, Monniaux M, Wu MF, et al. LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. Dev Cell. 2011; 20 (4):430–43. doi: 10.1016/j.devcel.2011.03.019 PMID: 21497757.

39. Benlloch R, Kim MC, Sayou C, Thevenon E, Parcy F, Nilsson O. Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of APETALA1. The Plant journal: for cell and molecular biology. 2011; 67(6):1094–102. doi: 10.1111/j.1365-313X.2011.04660.x PMID: 21623976.

40. Wagner D, Sablowski RW, Meyerowitz EM. Transcriptional activation of APETALA1 by LEAFY. Science. 1999; 285(5427):582–4. Epub 1999/07/27. PMID: 10417387.

41. Fourquin C, Ferrandiz C. Functional analyses of AGAMOUS family members in Nicotiana benthamiana clarify the evolution of early and late roles of C-function genes in eudicots. The Plant journal: for cell and molecular biology. 2012; 71(6):990–1001. doi: 10.1111/j.1365-313X.2012.05046.x PMID: 22563981.

42. Serwatowska J, Roque E, Gómez-Mena C, Constantin GD, Wen J, Mysore KS, et al. Two euAGAMOUS genes control the C-function in Medicago truncatula. PloS one. 2014;Accepted((In Press)).

43. Berg OG, von Hippel PH. Selection of DNA binding sites by regulatory proteins: Statistical-mechanical theory and application to operators and promoters. Journal of Molecular Biology. 1987; 193(4):723–43. http://dx.doi.org/10.1016/0022-2836(87)90354-8. PMID: 3612791

44. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic acids research. 2012; 40(Database issue):D1178–86. doi: 10.1093/nar/gkr944 PMID: 22110026; PubMed Central PMCID: PMC3245001.