### SOFTWARE



# JCoDA: a tool for detecting evolutionary selection

Steven N Steinway<sup>1</sup>, Ruth Dannenfelser<sup>1</sup>, Christopher D Laucius<sup>1</sup>, James E Hayes<sup>2</sup> and Sudhir Nayak<sup>\*1</sup>

#### Abstract

**Background:** The incorporation of annotated sequence information from multiple related species in commonly used databases (Ensembl, Flybase, Saccharomyces Genome Database, Wormbase, etc.) has increased dramatically over the last few years. This influx of information has provided a considerable amount of raw material for evaluation of evolutionary relationships. To aid in the process, we have developed JCoDA (Java Codon Delimited Alignment) as a simple-to-use visualization tool for the detection of site specific and regional positive/negative evolutionary selection amongst homologous coding sequences.

**Results:** JCoDA accepts user-inputted unaligned or pre-aligned coding sequences, performs a codon-delimited alignment using ClustalW, and determines the dN/dS calculations using PAML (Phylogenetic Analysis Using Maximum Likelihood, yn00 and codeml) in order to identify regions and sites under evolutionary selection. The JCoDA package includes a graphical interface for Phylip (Phylogeny Inference Package) to generate phylogenetic trees, manages formatting of all required file types, and streamlines passage of information between underlying programs. The raw data are output to user configurable graphs with sliding window options for straightforward visualization of pairwise or gene family comparisons. Additionally, codon-delimited alignments are output in a variety of common formats and all dN/dS calculations can be output in comma-separated value (CSV) format for downstream analysis. To illustrate the types of analyses that are facilitated by JCoDA, we have taken advantage of the well studied sex determination pathway in nematodes as well as the extensive sequence information available to identify genes under positive selection, examples of regional positive selection, and differences in selection based on the role of genes in the sex determination pathway.

**Conclusions:** JCoDA is a configurable, open source, user-friendly visualization tool for performing evolutionary analysis on homologous coding sequences. JCoDA can be used to rapidly screen for genes and regions of genes under selection using PAML. It can be freely downloaded at <u>http://www.tcnj.edu/~nayaklab/jcoda</u>.

#### Background

The first step in the assessment of evolutionary relationships between related sequences is the generation of pairwise or multiple sequence alignments (MSAs). Over the last two decades several algorithms have been developed to generate rapid yet accurate sequence alignments for subsequent analysis [1]. A commonly used program, ClustalW, generates MSAs of DNA or amino acids by constructing a branched guide tree from pairwise alignments [2]. More recent progressive alignment methods, such as T-COFFEE, have improved the accuracy of ClustalW by combining information from local and global alignments [3]. Other methods of sequence alignment, such as iterative, all-against-all, and hybrid approaches have also been shown to improve the accuracy of ClustalW, although some necessitate significant increases in computational power [4-7]. Regardless of the method, when DNA is aligned it is done so in a manner that arranges sequences to minimize gaps and mismatches to achieve a maximal score based on sequence identity and similarity.

Given that codon triplets are considered the unit of coding sequence evolution, DNA alignments that do not constrain codons are likely to misrepresent the encoded information [8]. For example, to meet the optimality criteria used in the alignment of DNA sequences, single gaps are frequently inserted and thus distort the reading frame (Figure 1, left). Essentially, for coding sequences, an "optimal" alignment of DNA may ignore the rules that govern its translation into protein. As a result, the evolu-



© 2010 Steinway et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons. Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

<sup>\*</sup> Correspondence: nayak@tcnj.edu

<sup>&</sup>lt;sup>1</sup> Department of Biology, The College of New Jersey, 2000 Pennington Road, Ewing, NJ 08628, USA

Full list of author information is available at the end of the article



tionary constraints placed on the protein product are lost in the analysis. A straightforward solution to the problem would be to perform a codon-based alignment that does not allow the partition of codons (Figure 1, right). The codon-based alignment can then be used to detect adaptive molecular evolution or purifying selection by estimating the number of non-synonymous and synonymous substitutions (dN/dS). In general, the aligned sequences are screened for dN/dS ratios of >1 (adaptive) or dN/dS ratios of < 1 (purifying).

There are currently several informatics tools freely available that calculate dN/dS ratios to measure evolutionary selection or generate codon-delimited alignments. The online program OCPAT is able to generate codon-delimited of alignments from human gene IDs and their putative orthologs from other vertebrates; however, OCPAT is not able to calculate dN/dS scores [9]. Most of the programs that calculate dN/dS require a codondelimited alignment. For example, the programs SNAP (Synonymous Non-synonymous Analysis Program) and WINA (Window Analysis) employ user provided alignments to calculate substitution rates, where WINA also allows for the use of sliding window analysis [10-12]. SWAKK (Sliding Window Analysis of Ka and Ks) uses pairwise sequence alignments, sliding windows, and structural alignment to identify regions of positive selection [13]. DNaSP (DNA Sequence Polymorphism) allows for the detection of diversifying selection by measuring DNA polymorphisms and also uses sliding window analysis [14].

In an effort to simplify the process, PAL2NAL (v12) takes pre-aligned protein and the corresponding unaligned coding DNA to generate codon-delimited alignments. While useful, PAL2NAL is constrained to pairwise dN/dS analysis, does not include graphing, and does not allow the user to edit the codeml control file [15]. The popular program PAML (Phylogenetic Analysis by Maximum Likelihood) is able to perform both pairwise and site-based dN/dS calculations. However, it does

not have a GUI (graphical user interface) and visualization of the output data can be cumbersome. In addition, it requires the input of pre-aligned sequences [8]. More recently, IDEA (Interactive Display For Evolutionary Analyses) has implemented a GUI for both PAML and Phylip (Phylogeny Inference Package) and is amenable for high throughput genome-wide analysis [4]. While IDEA is a powerful tool, it does not run on the Windows operating system, requires the separate installation of several support programs, uses multiple languages, generates graphs that are difficult to configure, and also requires a codon-delimited alignment.

We have designed JCoDA to be an easy-to-use tool that integrates several common functions required for the detection of evolutionary selection in coding sequences. Specifically, the steps that are now controlled from the JCoDA interface: generation of codon-delimited alignments, generation of phylogenetic trees, estimation of site and regional dN/dS scores under multiple models of substitution, and the generation of user configurable graphical output. JCoDA only requires unaligned coding sequences (CDS) in FASTA format and it takes advantage of the freely available BioJava framework [16], ClustalW [17], Phylip [18], and PAML [8] to identify positive/negative evolutionary selection. The visualization options in JCoDA include a variety of common alignment formats, and user configurable scalable graphs with sliding window options for pairwise or gene family comparisons. To demonstrate the types of analyses facilitated by JCoDA, we have performed an analysis of 25 sex determination genes in nematodes and identified genes under positive selection, regional positive selection, and differences in selective pressure likely due to functional constraints.

#### Implementation

JCoDA was designed to assist in performing common operations associated with evolutionary analysis. It coordinates the passage of information from the initial alignments (ClustalW version 2.0) to the calculation of dN/dS (yn00 or codeml, version 4.1) to the visualization of output. For site-based analysis of selection we have packaged JCoDA with a graphical interface for Phylip (version 3.6) that allows for the generation of phylogenetic trees. JCoDA is written entirely in Java, which allows for the addition of supplementary modules that offer additional functionality. To allow for easy installation, JCoDA is packaged with all of the required components (ClustalW, Phylip, and PAML) and can be installed on any computer with a Windows operating system (or virtual machine) and Java Runtime Environment 6 (JRE 6) (if Java 1.6 has been installed). For ease of operation, installation of Java Developer Kit 6 bundled with NetBeans (JDK, which includes JRE 6) is recommended http://java.sun.com/ javase/downloads/index.jsp.

JCoDA provides two options for calculation of dN/dS scores: regional pairwise calculation via sliding window or a site-specific calculation. Pairwise dN/dS performs the calculation between any user-selected sequences from a list of all possible comparisons presented in the GUI. For sliding window pairwise calculations, the size of the window, jump, and substitution models are configurable via drop down menus. Once the selected sequence comparisons are submitted, JCoDA parses through them by window, converts them to Phylip format, and feeds each window to PAML (yn00) suite to calculate dN/dS. JCoDA does this iteratively over all windows until the end of the selected sequence pair. The potential benefit of the sliding window option is that it can be performed very quickly and it is able to extract information about regional selection. However, it is important to note that methods that incorporate sliding windows have been demonstrated to be prone to artifact arising due to resampling and illustrate the importance of incorporating site-based methods and the likelihood ratio test in sequence analysis [19].

JCoDA takes advantage of the codeml executable included in the PAML package to implement site-specific dN/dS calculation. Similar to pairwise comparisons, JCoDA converts all inputted sequences to Phylip format and feeds them to codeml. In order for a site-based dN/ dS calculation to be performed, the user must provide a tree file and set its path for use in the GUI. To allow for this operation, the JCoDA package includes a Java based graphical interface for the Phylip package [18], http:// evolution.genetics.washington.edu/phylip.html). The Phylip Graphical Interface (PGI) allows for the generation of trees using neighbor-joining, parsimony, or maximum likelihood based methods of either DNA or protein. Regardless of the source of the tree file, JCoDA will accept any tree in Phylip format.

JCoDA implements M7 (fit to a beta distribution, dN/dS > 1 disallowed) -vs- M8 (fit to a beta distribution, dN/dS > 1 allowed) and M1a (nearly neutral) -vs- M2a (positive selection) in conjunction with the likelihood ratio test to check for evidence of positive selection [20]. To maintain JCoDA's flexibility, the user is given the option to access the complete codeml control file as a selectable advanced options tab to vary other parameters.

#### **Results and Discussion**

#### Input requirements and alignments

The JCoDA user interface, input requirements, and installation have been designed to be easier to use while retaining the underlying power of the codeml and yn00 programs from the PAML package (Additional File 1, Additional File 2, Additional File 3, Additional File 4 and Additional File 5). For example, we have simplified the input requirements to the coding sequences (CDS) in FASTA format. There were two primary reasons for this streamlined approach. First, CDS can be readily batch retrieved from NCBI or organism specific databases such as Wormbase (WormMart) [20]. Second, the risk of mismatch between DNA and protein sequences is eliminated by directly translating the CDS input by the user.

For unaligned CDS, the sequences are translated by the BioJava framework [16] and the proteins are passed to ClustalW for alignment. JCoDA generates a codondelimited alignment using the protein alignment from ClustalW as a guide to prevent the interruption of codon triplets, (Figure 1). JCoDA also accepts pre-aligned protein sequences in FASTA format paired with corresponding unaligned CDS in the same format to allow for the use of sequences aligned in another program or modified by hand. When provided with prealigned protein sequences and corresponding CDS, JCoDA simply circumvents ClustalW and performs the codon-delimited alignment directly.

#### Visualization and output

Navigation in JCoDA has been designed around the use of a tabbed GUI that allows for shuffling between graphs and sequence alignments (Figure 2). We have implemented JFreeChart <u>http://www.jfree.org</u> to generate robust graphs of dN/dS scores by both sliding window and site-specific methods. The graphs are extensively customizable, dynamically scaled, and can be saved as PNG (Portable Network Graphics) files (Figure 3 and 4A). The user can export protein and DNA alignment files in ClustalW, Phylip (v3.2 and v4), and hybrid protein/DNA codon-delimited formats using the file menu. The dN/dS scores from all models selected, including p values, can be exported as common separated values (CSV) files allowing for further analysis in database, spreadsheet, and graphing programs (Figure 4B).

#### Benchmarking

The use of JCoDA interface does not add significantly to the time need required to run ClustalW, Phylip, or PAML. Any time costs incurred are more than ameliorated by the integration of multiple tasks. For example, the user does not have to reformat sequences to shuttle them from one program to another or use an additional program to obtain graphical representations of the data. To test the ability of JCoDA to recover signatures of directional selection, we used multiple datasets including TRIM5α (hominids + OWMs) [21], nef gene of HIV-1 (pairwise and ML) [22], lysozyme [23], and a subset of HIV1 protease and reverse transcriptase sequences from the HIV positive selection mutation database [24]. We have included an analysis of the sex determination genes in nematodes to illustrate the utility of JCoDA in identification of directional selection.

Alignment Options:		Input cDNA Output CoDA Format Output Aln - Prot Output Aln - DNA	
() CLUETAL (Sample)		apple come Courrenties Courrenties Courses and Courses and	
() Fre-Aligned (Susple)		Codon Delimited Alignment:	
Submit Reset Help		CLUSIAL multiple sequence alignment	
En la seria		Ce_gld-1 atgccgtcgtgcsccactocsacttscggtgtttcgacccsactsgasagccasagctct	
Traphing Options:		Cja_gld- atgccgtcgtgtaccacgccaacattcggggcatctcaattagagatcgggt Cre_gld- atgccgtcgtgcaccacgccaacattcggggcatctcaattagagatcggggt	
Fliding Window:		Cbr_gld- atgccgtogtgcaccacoccaspcttoggtgtgtoggcacaggagaacccaagogtc	
Window: 200 Jump:	13 💌	Chn_gld- atgccgtcctgcaccacccccccctacqgtgtaccggcccaactcgacagccaaagctca	
B/d3 Substitution Model: (7)		Ce ald-1 gestconcetclogticgtcontetgeconcesogtctctggstggsgesstagtcos	
Her-Hojobber (1966)		Cja_gld- gattogcogagtogtgagtotgtcatgactcogacttotttagatgacgacacgcca	
		Cre_gld- gagtotcogtcacgttoctcgattctpacgcctacgtcgttggaagatgagaccagocct	
alculations: To	Be Graphed:	Lor_gis- gattcaccattcogttogttogttogtcoccascattcattcgaogatgagaccagtcoa Cbn gld- gastcgcogtctogttogttogttogtcoccascattcattcgaoggagacsacagccog	
> Ce.	gid-i ve Cja_g A		
< Q.	glé-1 ve Ger_gin	Ce_gid-1 cggaagagattccccatostcgacaatgttccagoggatogttggccgtca	
Dx Dx	gld-1 ve Can_gl	Cre_gld- cgasagsgatttccgctgattgagacgsatatcagcgcttcggatcgttggccgccagct	
29 C14	gld- rs Car g	Cbr_gld- cggaagagtttccagattttcgagtccagtgtgagogcggatogttggccagcg	
		.nn_gig- ageaagogettocogatggectogtogettocectctoggetogotgggccgca	
Course of Laboration Strength and the	1	Ce_gld-1 actoggogagacggatggagctctgttagagcgocaccagctogtctgacgctatcg	
oraph Silding window dB/dS		Cja_gld- tatcggcgagacggatggagctetttgcggggcgccacetecagetcgcetetegetgtee	
Calculate d0/d5 by site:		Chr_gld- coacggogagatggctggagtoagtggagaccaccaccagctogtctctcccccas	
house Model: (7)	-1	Cbn_gld- coscgsogggatggatggsgctocgttogtgcsccsccsccsgctogtctctcgctccac	
hoose Tree File:		Ce gld-1 acgascascagecacatostgtocccgatttcgsgcgcttacagtcaascacctasttcg	
	Bauwas.	Cja_gld- actaaccacaagaactttatgtctccgatttcgagtgcatacagtcaaacgccaaattcg	
Graph d0/dE by Eites	Build a Tree	Cre_gld- accasocacagasgcatgctttopccastatccagogcgtacagtcagacaccgaactcg	
TIRA' ALCON	2 Militas mathods remitred	Chn gld- accasciacamasacricatorccastatorccastatorccastatorccastcr	
Advances ("Hote: Minimum of www.teng.udu/-maysklab/joods.t Tatus: 	1151 1 +000+1+Ful1		
Anvances ("Poots: Ministen o mov.comj.etu/-nayaklah/joods.i Poster-delimited aligneen 3 PGL	neni e rescentrăți e		
JARVARGES (*POCE: KILLIGHE C POCE: HUJ-nageklah/grode.) Tatus: Deserfelented alignee J PGL Govert Maxmun Ließhod Tree	todi • recentrăți   •		
Antranses ( 1905s Initialian of	tani + reconstruit	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY	
Antrances (*Note Attransmo) overset, sub/good i tatus: Diserdelisited eligness pcc fect Convert Maximum LiseRood Tree Enable Bootstrapping? Ves •	teni + reconstruit	Input Window (PHYLIP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Import Converted File	
Antranses (*pose initialized) or cend; emb/mageblack/good i tratue: Deter-delimited alignee g PGL Get Convert Nammu Lesihood Tree Enable Bootstrapping? Ves N Hove many reploates?	seal * recommendations Consense Options Consense type [liaprov nule (extended)] •	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Input Converted File 5 488 C.e., gid-1 MPSCTTPTYG VSTOLESOSS ESPRESSVMT PTSLDDDT-PRX/FPTLETGDRWP-M 0.0, gid-1 MPSCTTPTTGDRWP-M	
Antranses (*1005 : 10.015m of more cent, entry angehilde/jords : Tratus : Deter-deliasted alignee 3 PGI 198 Convert : Naximum Likelihood Tree Enable Bootstrapping? Ves	e reserve fui (	Input Window (PHYLP 3.88 FORMAT-INTERLEAVED SEQUENCES ONLY) Import Converted File 5 488 Oc. gid-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSVMT PTSLDODNSP RKREPPIDHVPADRWP-S Cjagid- MPSCTTPTFG AS-OLEBOSS DESPSRSSVMT PTSLDODT-P RKREPPLETN CASORWEPA CPLgid- MPSCTTPTFG VTSOLETCMA ESPSRSSLT PTSLEDETSP RKREPPLETN ISASORWEPA	
Antrades (*1005 filling) Antrades (*1005 filling) Platus (Index) Platus (In	consense Options Consense Options Consense type (liaporty nic contend) • Praction ():	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Import Converted File 5 488 Co., gld-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSVMT PTSLDGDNSP RKRFPIDNVPADRWP-S Cip., gld- MPSCTTPTFG A3-OLEIGSG DSPSRESVMT PTSLDGDT-P RXRFPIDNVPADRWP-S Com, gd MPSCTTPTFG VSADLETONA ESPSRSSLT PTSLDGETSP RKRFPILETN ISASDRWPPA Com, gd MPSCTTPTFG VSADLETONA ESPSRSSLT PTSLDGETSP RKRFPI-MDSS IPL SORWAPA	
Artvaces (*8055 *10.158m 0) or cend; etb./apak3.ab/j.dod . tratus : Deter-deliaited elignese (************************************	ten1 * recommendations Consense Options Consense type [itacrit rule contended] • Fraction () Treat trees as noted? [ito] •	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Import Converted File 5 488 Co., gid-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSVMT PTSLDCDIVSP RKRFPIDNVPADRWP-S Co., gid-1 MPSCTTPTFG AS-OLEDOSC DSPSRESVMT PTSLDCDIT-P RKRFPILETGDRWP-M Co., gid MPSCTTPTFG VSADLEDASC DSPSRSSLT PTSLDCETSP RKRFPILETI SASDRWPPA Cor_, gid MPSCTTPTFG VSADLEDASC SESPSRSSLT PTSLDCETSP RKRFP-MIDSS IPLSDRWA	
Arvanese (*2006 *10.15.88 00 00 00 00 00 00 00 00 00 00 00 00 0	teni * reconcertfull  Consense Options Consense type Islantv rule (calended)   • Fraction (n) Treat trees as rooted?  To Search for best tree?*	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) INSOLTCONVITED FROM 5 488 Ce_gid-1 MPSCTTPTYG VSTOLESOSS ESPSRSSVMT PTSLDCDNSP RXRFPIDIHVPADRWP-5 Cor_gid- MPSCTTPTYG VSTOLESOSS ESPSRSSVMT PTSLDCDNSP RXRFPIDILESV SA-ORMVP-A Cor_gid- MPSCTTPTYG VSOLETORA ESPSRSSLT PTSLDCDTSP RXRSFPLETN ISASORWPA Cor_gid- MPSCTTPTYG VSOLETORA ESPSRSSLT PTSLDCDTSP RXRSFPLETN ISASORWPA Cor_gid- MPSCTTPTYG VSOLETORA ESPSRSSLT PTSLDCDTSP RXRSFPLESS VSA-ORMVP-A Con_gid- MPSCTTPTYG VFADLDSQSS ESPSRSSLT PTSLDCQNSP RXRFP-MOSS IPLSDRWA- TRRDCWSSVR APPPARLTLS TNINRHIMSPI SSAYSOTPHS LLSPAMIFYK RRSFSPTLP VRRDCWSSVR APPPARLSLS TNINRHIMSPI SSAYSOTPHS LLSPAMIFYK RRSFSPTLP	
Artvanese (*2005 * 10.01580 10.000) Processing (*2005 * 10.01580 10.000) Processing (*2005 * 10.000)	Consense Options Consense Options Consense type (Raphy vice (addender)) • Fraction (t) Treat trees as rooted? No Search for best bee?* Yea •	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Input Converted File 5 488 Cal.gld-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSWIT PTSLDGDNSP RKRFPIDNVPADRWP-S Cal.gld-MPSCTTPTFG AS-OLEBOSS DSPSRSSWIT PTSLDGDT-P RKXFPIDNVPADRWP-M Col.gld-MPSCTTPTFG VTSOLETONA ESPSRSSLT PTSLDETSP RKXFPIDETN ISASDRWPA Col.gld-MPSCTTPTFG VTSOLETONA ESPSRSSLT PTSLDGDTSP RKXFPIDETN SASDRWPA Col.gld-MPSCTTPTFG VTSOLETONA ESPSRSSLT PTSLDGTSP RKXFPIDETN SASDRWPA Col.gld-MPSCTTPTYG VAOLENGSS ESPSRSSLT PTSLDGDNSP RKRFP-MDSS IPLSORWA- TRRDGWSSKR APPPARLTLS TNARHUSPI SSAYSOTTPNS LLSPAMIPMK SRSSPSPTLP YRRDGWSSKR APPPARLSLS TNAHOLMSPI SSAYSOTTPNS LLSPAMIPMK TRSFSPTLP PRRDGWSSKR APPPARLSLS TNAHOLMSPI SSAYSOTTPNS LLSPAMIPMK TRSFSPTLP	
Artvaces (*2006 *20.016 *0.016	e recentful e recentful Consense Options Consense Options Consense type [lagorir rule (addredd)] • Praction () Praction () Fraction () Search for best free?* Vas lengths from user trees?	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Insect Converted File 5 488 Ce.gld-1 IMPSCTTPTYG VSTOLEBOGS ESPSRSSVMT PTSLDGDNSP RKRFPIDNVPADRWP-S Citgid- MPSCTTPTYG AS-OLEIGSG DSPSRSSVMT PTSLDDDT-R RKKFPILDNVPADRWP-S Citgid- MPSCTTPTYG VSTOLEDGYS ESPSRSST PTSLEDETSP RKRFPILETN GASORWPPA Cotgid- MPSCTTPTYG VSADLEIGAS DSPSRSSLT PTSLDDETSP RKRFPILEN IGASORWPPA Cotgid- MPSCTTPTYG VSADLEIGAS PSRSSLT PTSLDDGTSP RKRFPILDN ISS.DRWA- Cotgid- MPSCTTPTYG VSADLEIGAS PSRSSSLT PTSLDDGTSP RKRFPILDN ISS.DRWA- TRRDGWSSVR APPPARLTLS TNIRHIMSPI SSAYSOTPNS LLSPAMFNAK RRSIFSPTLP PRRDGWSSVR APPPARLSLST THHKIPKISPI SSAYSOTPNS LLSPAMFNAK RRSIFSPTLP PRRDGWSSVR APPPARLSLST THHKIPKINSPI SSAYSOTPNS LLSPAMFNAK RRSIFSPTLP	
Artvandes (*8055 *8101588 00) Artvandes (*8055 *810158 00) Post Partie Convert Nassman L&Rithood Tree Enable Bootstrapping? Ves V How many replicates? Promit Options Select probability model James Taylor Thornton Ratv variation among sites constant rath of cha. V Coefficient of var # of categories	tan1	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY Input Converted File 5 488 Co., gld-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSVMT PTSLDCDN-SP RKRFPHDNVPADRWP-S Co., gld-MPSCTTPTFG AS-OLEIGSG DSPSRESVMT PTSLDCDT-P RKXFPLETGDRWP-M Co., gld-MPSCTTPTFG VSDLETGAK ESPSRSSLT PTSLDCETSP RKRFPHDSS INSCRWPPA Com, gld-MPSCTTPTG VSDLESS ESPSRSSSLT PTSLLSPMTKK KRSFSPTLP PRRDCWSSWR APPPARLSLS TIMENIUSPI SSAYSOTTPKS LLSPMTKHK KRSFSPTLP PRRDCWSSWR APPPARLSLH HTMHRIMSPI SSAYSOTTPKS LLSPMTKHK KRSFSPTLP PRRDCWSSWR APPPARLSLH HTMHRMMSPI SSAYSOTTPKS LLSPMTKHK KRSFSPTLP PRRDCWSSWR APPPARLSLH HTMHRMMSPI SSAYSOTTPKS LLSPMTKHK KRSFSPTLP	
Artvaces (*8055 *10.15.80 0) Artvaces (*8055 *10.15.80 0) (*805 - 10.15.80 0) (*10.15.80 0)	Consense Options Consense Options Consense type Illatority rule (cadender) v Praction () Treat trees as noted? No v Search for best tree?* Vestor () Use lengths from user trees? Rate(s) for each category	Input Window (PHYLP 2.68 FORMAT-INTERLEAVED SEQUENCES ONLY) INSURT Converted File 5.488 C. glid-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSVMT PTSLDCDN-SP RKRFPIDNVPADRWP-S C. glid-1 MPSCTTPTFG AS-OLEDGS DSPSRESVMT PTSLDCDT-P RKRFPIDNVPADRWP-S C. glid-1 MPSCTTPTFG VSDLETGAR ESPSRSSLT PTSLDCETSP RKRFPIDNVPADRWP-S C. glid-1 MPSCTTPTFG VSDLETGAR ESPSRSSLT PTSLDCETSP RKRFP-MDSS ILSSRWPPA C. glid-1 MPSCTTPTFG VSDLETGAR ESPSRSSLT PTSLDCETSP RKRFP-MDSS ILSSRWPPA C. glid-1 MPSCTTPTFG VSDLETGAR ESPSRSSLT PTSLDCETSP RKRFP-MDSS ILSSRWPPA TRRDGWSSVR APPPARLTLS TNINRHIMSPI SSAYSOTTPNS LLSPTMFHK SRSIFSPTLP PRRDGWSSVR APPPARLSLS TNI-NAFIMSPI SSAYSOTTPNS LLSPTMFHK SRSIFSPTLP PRRDGWSSVR APPPARLSLH TNI-RISMLSPI SSAYSOTTPNS LLSPTMFHK TRRSFSPTLP PRRDGWSSVR APPPARLSLH TNI-RISMLSPI SSAYSOTTPNS LLSPTMFHK TRRSFSPTLP PRRDGWSSVR APPPARLSLH TNI-RISMLSPI SSAYSOTTPNS LLSPTMFHK SRSIFSPTLP PRRDGWSSVR APPPARLSLH TNI-RISMLSPI SSAYSOTTPNS LLSPTMFHK TRRSFSPTLP PRRDGWSSVR APPPARLSLH TNI-RISMLSPI SSAYSOTTPNS LLSPTMFHK SRSIFSPTLP PRRDGWSSVR APPPARLSLH TNI-RISMLSPI SSAYSOTTPNS LLSPTMFHK SRSIFSPTLP PRRDGWSSVR APPPARLSLH TO TNI-RISMLSPI SSAYSOTTPNS LLSPTMFHK SRSIFSPTLP PRRDGWSSVR APPPARLSPARTEPPEVER TVEVLOUWLSPNK MHYRMFHK SRSIFSPTLP PRRDGWSSVR APPPARLSPARTEPPEVER TVEVLOUWLSPNK MHYRMFHK SRSIFSPTLP PRRDGWSSVR APPPARLSPARTEPPEVER TVEVLOUWLSPNK MHYRMFHK SRSIFSPTLP	
Artvances (*8055 * 10.1.5.00 *	Consense Options Consense Options Consense type Itacirity rule (celended) v Fraction (i) Treat trees as rooted? No Search for best hee?" Vea Search for best hee?" Vea Use lengths from user trees? Rate(s) for each category One category of sales?"	Input Window (PHYLP 3.88 FORMAT-INTERLEAVED SEQUENCES ONLY) Import Converted File 5 488 Oc. glid-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSVMT PTSLDODNSP RKREPPIDHVPADRWP-S Oc. glid-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSVMT PTSLDODT-P RKREPPIDHVPADRWP-S Con_glid- MPSCTTPTYG VSADLEDGS OSPSRESMIT PTSLDODT-P RKREPPIDH-SADRWP-A Con_glid- MPSCTTPTYG VSADLENGS OSPSRESMIT PTSLDOTSP RKREPPIDH ISASDRWPA Con_glid- MPSCTTPTYG VSADLENGS ESPSRSSLT PTSLDOTSP RKREPPIDEN ISASDRWPA Con_glid- MPSCTTPTYG VADLEDGSS ESPSRSSLT PTSLDOETSP RKREPPIDEN ISASDRWPA Con_glid- MPSCTTPTYG VADLEDGSS ESPSRSSLT PTSLDOETSP RKREPPIDES INSORWA- TRRDOWSSLR APPPARLSLT THINREIMSP ISANSOTTMS LLSPTMFHKK RREISPTLP PRRDOWSSLR APPPARLSLT THINREIMSP ISANSOTTMS LLSPTMFHKK SRBIESPTLP PRRDOWSSIR APPPARLSLT THINREIMSP ISANSOTTMS LLSPTMFHKK SRBIESPTLP PRRDOWSSIR APPPARLSLT THINREIMSP ISANSOTTMS LLSPTMFHKK SRBIESPTLP PRRDOWSSIR APPPARLSLH THINREINSP ISANSOTTMS LLSPTMFHKK SRBIESPTLP PRRDOWSSIR APPPARLSLH THINREINSP ISANSOTTMS LLSPTMFHKK SRBIESPTLP PRRDOWSSIR APPPARLSLH THINREITEPTEVELAUKENKHMINNER ISANSOTTHMS LLSPTMFHKK SRBIESPTLP	
Artvaces (*socs transmon) Artvaces (*socs transmon) (*socs trans	ten1 a reconstruct Consense Options Consense Options Consenses type Illaority rule (extended)   • Fraction (*) Treat these as rooted? No • Search for best thee?* Yes Use lengths from user thees? Rate(s) for each category Consensery of ables?* Yes Ves •	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) INSUE TO AS A DESCRIPTION OF THE ASSAULT ANTERLEAVED SEQUENCES ONLY) INSUE TO ASSAULT ANTERLEAVED SEQUENCES ONLY AND	
Antrances (*8055 *10.15.800 ) or.end; etd., rangeblack/good ; tatus: Describelingted the tatus: Describelingted the ford: Bet Enable Bootstrapping? Ves V How may replates? Prom Options Select probability model Janes-Tavlor-Thornton V Rate variation among sites constant rate of dat V Conferent of www # of categores Conferent of www # of categores Fraction of invariant abs? Speeder to troopler analysis? Ves V	tan1 a more marked options Conserve Options Conserve type [liaborb rule rodended] ♥ Fraction (); Time trees as rooted? ♥ Fraction (); Search for best tree?* ♥ Vices ♥ Use lengths from user trees? Use lengths from user trees? Rate(s) for each category One category of sates?* ♥ Probabilities for each category	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Insect Converted File 5 488 C.g.gld-1 IIPSCTTPTFO AS-OLEIGSS DSPSRESMIT PTSLDDDT-R RKKFPIIDNVPADRWP-S Citgd- MPSCTTPTFO AS-OLEIGSS DSPSRESMIT PTSLDDDT-R RKKFPIIDNVPADRWP-S Citgd- INPSCTTPTFO VSADLEIGNS DSPSRESMIT PTSLDDETSP RKRFPHEN ISS/SOKWPPA Citgd- INPSCTTPTFO VSADLEIGNS DSPSRESMIT PTSLDETSP RKSFDLESS VISADRWPA RKRDGWSSVR APPPARLISLT TINKRHIMSPI SSAYSOTPNS LLSPAMFNK SRSIFSPTLP PRRDGWSSVR APPPARLISLH TINKRHIMSPI SSAYSOTPNS LLSPAMFNK SRSIFSPTLP PRRDGWSS BDKS-LSP TATEPREVENT TYCNLAUK EXONLTLPH MENNVERLLD STPMISTOKS BDKS-LSP TATEPREVENT VEYLAUK EKONLTLPH MENNVERLLD ATPLSYGKSS BDKS-LSP TITEPREVENT VEYLAUK EKONLTLPH MENNVERLLD ATPLSYGKSS TEKS-LSP TITEPREVENT VEYLAUK EKONLTLPH MENNVERLLD	
Artvardes (*8055 *81.11.51.000 ) or cend; self. Anayaki Lariy dood . Tabila: Deter-del larived a Lagness tabila: Deter-del larived a Lagness tabila: Deter-del larived a Lagness table Bootstrapping? Yes How may replicates? Promit Options Select probability model James-Tavlor-Thornton Promit Options Select probability model James-Tavlor-Thornton Rate variation among alles constant rate of cha. * Coofficient of var # of categories Fraction of rowards taba? Speeder but rougher anayais? Yes Speeder but rougher anayais? Yes Coofficient of var # of categories Coofficient but rougher anayais? Yes	ten1	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Input Converted File 5 488 Co., glid-1 MPSCTTPTYG VSTOLESOSS ESPSRSSVMT PTSLDCDN-SP RKRFPHDNVPADRWP-S Cit., glid-MPSCTTPTFG AS-OLEIGSG DSPSRESVMT PTSLDCDT-P RKKPFLETGDRWP-M Cot., glid-MPSCTTPTFG VSLDLETMA ESPSRSLT PTSLDCETSP RKRFPHDNVPADRWP-S Cot., glid-MPSCTTPTFG VSLDLETMA ESPSRSLT PTSLDCETSP RKRFPHDSS IPLSORWPPA Cot., glid-MPSCTTPTFG VSLDLETMA ESPSRSLT PTSLDCETSP RKRFPHDSS IPLSORWPPA Cot., glid-MPSCTTPTFG VSLDLETMA ESPSRSLT PTSLDCOTSP RKSFOLESS VSLDRWA- Cot., glid-MPSCTTPTFG VSLDFPARLSLS THHON MSPI SSAYSOTTPKS LLSPAMINEY SRSFSPTLP PRRDOWSSIR APPPARLSLS THHON MSPI SSAYSOTTPKS LLSPAMINEY KSRSFSPTLP PRRDOWSSIR APPPARLSLY THANKINGSI SSAYSOTTPKS LLSPAMINEY KSRSFSPTLP PRRDOWSSIR APPPARLSLY THEPKEVER TVEYLADLYK EKKHLTLFPH MENNYERLD STPMSYGKSS LDKSKILSP PATEPYEVER TVEYLADLYK EKKHLTLFPH MENNYERLD ATPLSYGKSS TEKS-LSP TTTEPYEVER TVEYLADLYK EKKHLTLFPH MENNYERLD DEIGRYRVAL FOTEFPRVEL PEPAGOMIST TEKVYDAUE PROVINCIE (DEPCHTAKO DEIGRYRVAL FOTEFPRVEL PEPAGOMIST TEKVYDAUE PROVINCIE (DEPCHTAKO	
Artvardes (*8055 *10.15.800 ) or cent; self, range & Larly good , i tatus : Drem-del lasted e Lagreen iter Convert : Nassean Likelhood Tree Enable Bootstrapping? Ves Hour many replantes? Promit Options Select probability model Janes -Twylor-Thermion · Rain variation among sites constant rain of cha. · Coefficient of var # of categores constant rain of cha. · Coefficient of var # of categores constant rain of cha. · Coefficient of var # of categores constant rain of cha. · Fraction of hour # of categores constant rain of cha. · Speeder but rougher analysis? Ves · Speeder but rougher analysis? He · ·	ten1   Consense Options  Consense Options  Consense type  Lisoror nucleatedori  Fraction ()  Fraction ()  Trait tees as rooted?  No  Search for best tee?*  Yes  Use lengts from user trees?  Rate(s) for each category  One category of falts?*  Yes  Disc integrate for each category  Search weighted?*  Iso  Search weigh	Input Window (PHYLP 3.88 FORMAT-INTERLEAVED SEQUENCES ONLY) Import Converted File 5.488 C.g., glo-1. MPSCTTPTYG VSTOLESOSS ESP3RSSVMT PTSLDCDN-SP RKRFPHDNVPADRWP-S CDL_306 MPSCTTPTFG AS-OLEICSG DSP3RESVMT PTSLDCDT-P RKRFPHDNVPADRWP-S CDL_306 MPSCTTPTFG VSDLETOAK ESP3RSSLT PTSLDCETSP RKRFPHDNVPADRWP-S CDL_306 MPSCTTPTFG VSDLETOAK ESP3RSSLT PTSLDCETSP RKRFP-MDSS ILSSRWPPA CDL_306 MPSCTTPTFG VSDLETOAK ESP3RSSLT PTSLDCETSP RKRFP-MDSS ILSSRWPPA TRRDGWSSKR APPPARLISLS TNIKRIMSPI SSAYSOTTPKS LLSPMIFNEK SRSIFSPTLP PRRDGWSSKR APPPARLISLH TNIHRIMSPI SSAYSOTTPKS LLSPMIFNEK SRSIFSPTLP PRRDGWSSKR APPPARLISL PSP TTTEPVEKA TVEYLADLK EXXHLTLPH MENNVERLLD STPMISYOKSS BUKSLFSP TITEPVEKA TVEYLADLK EXXHLTLPH MENNVERLLD ATPLSYCKSS TEKS-LFSP TITEPVEKA TVEYLADLK EXXHLTLPH MENNVERLLD DEIGRVRVAL, FOTEFPRVEL PEPAGOMIST TEKNYVPNOL (PDVIFVGRI LGPRGMTAKO DEIGRVRVAL, POTEFPRVEL PEPAGOMIST TEKNYVPNOL (PDVIFVGRI LGPRGMTAKO	
Artvances (*8055 *10.15.800 ) Artvances (*8055 *10.15.800 ) Tattus: Disker-deliaused eligneen itattus: Disker-deligneen itattus: Disker-deliaused eligneen itattus: Diske	Consense Options Consense Options Consense Options Consense type Itacrit vise contender) Fraction ()	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Import Converted File  5.488  C.g. gid-1 MPSCTTPTYG VSTOLEBOSS ESP3RSSVMT PTSLDCDNSP RKREPPIDNVPADRWP-S C.g. gid-1 MPSCTTPTFG AS-OLEDGS OSP3RESMT PTSLDCDT-P RKREPPIDNVPADRWP-S C.g. gid-1 MPSCTTPTFG VSDLETOAL ESP3RSSLT PTSLDEETSP RKREP-MOSS INSO-MWP-A Congid-1 MPSCTTPTFG VSDLEST NHKRHMSPI SSAVSOTTNS LLSPTMFHK RKREPSPTLP PRRDCWSSIR APPPARLSLST THHRIMEN SAVSOTTNS LLSPTMFHK RKREPSPTLP PRRDCWSSIR APPPARLSLST THHRIMEN SAVSOTTNS LLSPTMFHK RKREPSPTLP PRRDCWSSIR APPPARLSLH TVHINGPI SSAVSOTTNS LLSPTMFHK RKREPSPTLP PRRDCWSSIR APPPARLSP PTTEPVEAL TVETALLW EXXHLT.PFH MFNWRELLD STPMISTOKSS MOKS-LFSP TTTEPVEAL TVETALLW EXXHLT.PFH MFNWRELLD ATPLSYCKSS MOKS-LFSP TTTEPVEAL TVETALLW EXXHLT.PFH MFNWRELLD DEIGRVRVAL FOTEFPRVEL PEPAGOMIST TEKVYVNOV PYVVRVI LOPRGMTAXO DEIGRVRVAL FOTEFPRVEL PEPAGOMIST TEKVYVNOV PYVVRVI LOPRGMTAXO DEIGRVRVAL FOTEFPRVEL PEPAGOMIST TEKVYVNOV PYVVRVIG LOPRGMTAXO DEIGRVRVAL FOTEFPRVEL PEPAGOMIST TEKVYVPNUR PHYVVRVRUE DEPAGINATAXO DEIGRVRVAL FOTEFPRVEL PE	
Artvances (*8055 * 10.1.5.800 * 10.1.5.800 * 10.1.500 * 10.1 * 10	tenil  Conserve Options Conserve Options Conserve type Itaborty rule (celended)  Fraction (i) Treat trees as rooted? Fraction (ii) Search for best tree?* Vea  Use lengths from user trees? Use lengths from user trees? Rate(s), for each category One category of site?* Vea  Probabilities for each category Conserve of the site optimized of the site opti	Input Window (PHYLP 3.88 FORMAT-INTERLEAVED SEQUENCES ONLY)  Input Window (PHYLP 3.88 FORMAT-INTERLIAVED SEQUENCES ONLY)  Input Window (PHYLP 3.88 FORMAT-INTERLIAVED SEQUENCES ONLY)  Input Sequences of the PHYLP 3.88 FORMAT-INTERLIAVED SEQUENCES ONLY)  Input Sequences of the PHYLP 3.88 FORMAT-INTERLIAVED SEQUENCES ONLY)  Input Sequences of the PHYLP 3.88 FORMAT-INTERLIAVED SEQUENCES ONLY)  Input Sequences of the PHYLP 3.18 FORMATANA ON THE INTERLIAVED SET THE PHYLP A TIPE YALE A TYLY ALLY K EXAHLILIAPH MENIVERILLD  DEIGR/WRVAL FOTEPRYLE PEPAGOMIST TEXPYYPANE PHYLPAR ILSPRAIT-INAVE AND DEIGR/WRVAL FOTEPRYLE PEPAGOMIST TEXPYYPANE PHYLPAR ILSPRAIT-INAVE DEIGR/WRVAL FOTEPRYLE PEPAGOMIST TEXPYYPANE	
Antivances (*8005 *10.11580 0) Antivances (*8005 *10.11580 0) Poll Po	e movement fu's ( e movement fu's ( concernes type) Concernes type ( liabort rule (colorded)) Fraction (); Treat trees as rooted? Treat treat	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Insoft Converted File 5 488 C.g. gli-1 IMPSCTTPTTO 45-OLEGOS DEPSRESMIT PTSLDDDT-P RKXFPTLETG — DRWP-M Ctr., gli-MPSCTTPTTO 45-OLEGOS DEPSRESMIT PTSLDDDT-P RKXFPTLETG — DRWP-M Ctr., gli-MPSCTTPTTO 45-OLEGOS DEPSRESMIT PTSLDDDT-P RKXFPTLETG — DRWP-M Ctr., gli-MPSCTTPTTO 45-OLEGOS DEPSRESMIT PTSLDDETSP RKXFPTLETG — DRWP-M Ctr., gli-MPSCTTPTTO 45-OLEGOS DEPSRESMIT PTSLDDETSP RKXFPTLETG — DRWP-M Ctr., gli-MPSCTTPTTO 45-OLEGOS DEPSRESMIT PTSLDDETSP RKXFPTMETN ISS.DSRWPPA Ctr., gli-MPSCTTPTTO 45-OLEGOS DEPSRESMIT PTSLDCOMSP RKXFPTMDS IN 25.007WA Ctr., gli-MPSCTTPTTO 45-OLEGOS DEPSRESMIT PTSLDCOMSP RKXFPTMDS IN 25.007WA TRRDOWSSKR APPPARLSLS TNIKANEMSPI SSAYSOTTPKS LLSPAMFNeK TRSSFSPTLP PRRDOWSSKR APPPARLSLS TNIKANEMSPI SSAYSOTTPKS LLSPAMFNEK TRSSFSPTLP PRRDOWSSKR APPPARLSLS TNIKANEMSPI SSAYSOTTPKS LLSPAMFNEK TRSSFSPTLP PRRDOWSSKR APPPARLSL HT THRIANSPI SSAYSOTTPKS LLSPAMFNEK TRSSFSPTLP PRRDOWSSKR APPPARLSL HT THRIANSPI SSAYSOTTPKS LLSPAMFNEK SRRIESPTLP PRRDOWSSKR APPPARLSL HTREVEN TY VICAUL KENNALTLPPH MENNVERLLD STPMISTGKSS BDKS-LSP TTTEPPEVER TVEYLAQL KENNALTLPPH MENNVERLD DEIGRWRVAL FOTEPPRVEL PEPAGDMIST TEKVYVNKE ENNALTLPPH MENNVERLD DEIGRWRVAL FOTEPPRVEL PEPAGDMIST TEKVYVNKKE POVINFUGRI LGPRGITAKO DEIGRWRVAL FOTEPPRVEL PEPAGDMIST TEKVYVNKKE POVINFUGRI LGPRGITAKO DEIGRWRVAL FOTEPPRVEL PEPAGDMIST TEKVYVNKKE POVINFUGRI LGPRGITAKO DEIGRWRVAL FOTEPRVEL PEPAGDMIST TEKVYVNKKE POVINFUGRI LGPRGITAKO DEIGRWRVAL FOTEPRVEL PEPAGDMIST TEKVYVNKKE POVINFUGRI LGPRGITAKO DEIGRWRVAL FOTEPRVEL PEPAGDMIST TEKVYVNKKE POVINFUGRI LGPRGITAKO DEIGR	
Artvakes (*8054 *81115860 0) Artvakes (*8054 *1115860 ) Poll Po	e more and full in the second	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Insuft Converted File 5 488 C.g. gli-1 IMPSCTTPTFO AS-OLEGOS DEPSRESMIT PTSLDDDT-P RKXFPTLETG — DRWP-M Ctr., gli-MPSCTTPTFO VSTDLEBOSS ESPSRSSIT PTSLDDDT-P RKXFPTLETG — DRWP-M Ctr., gli-MPSCTTPTFO VSLDLETMA ESPSRSIT PTSLDDETSP RKRFPTLETG — DRWP-M Ctr., gli-MPSCTTPTFO VSLDLETMA ESPSRSIT PTSLDDETSP RKRFPTLETG — DRWP-M Ctr., gli-MPSCTTPTFO VSLDLETMA ESPSRSIT PTSLDDETSP RKRFPTLETG — DRWP-M Ctr., gli-MPSCTTPTFO VSLDLETMA ESPSRSIT PTSLDCOMPR RKRFPTLETG — DRWP-M Ctr., gli-MPSCTTPTFO VSLDLETMA ESPSRSIT PTSLDCOMPR RKRFPTLB VSRDOWSSUR PRODUMSSUR APPARLELS TINKINFINSPI SSAYSOTTPKS LLSPMFNHFV RSRSPTLP PRRDDWSSUR APPARLELS TINKINFINSPI SSAYSOTTPKS LLSPMFNHFV RSRSPTLP PRRDDWSSUR APPARLELS TINKINFINSPI SSAYSOTTPKS LLSPTMFNHFV RSRSPTLP PRRDDWSSUR APPARLES TITEPVEVA TVEVLAULKE EXXILTPH MENNVERLLD STPMISTGKVS LDSKSILFSP APTEPVEVA TVEVLAULKE EXXILTPH MENNVERLLD STPMISTGKVS LDSKSILFSP TTTEPVEVA TVEVLAULKE EXXILTPH MENNVERLLD DEIGRVRVAL FOTEFPRVEL PEPAGDMIST TEKVYVNE EXXILTPH MENNVERLLD DEIGRVRVAL FOTEFPRVEL PEPAGDMIST TEKVYVNENE YPDVIFVGRI LGPRGITAKO DEIGRVRVAL FOTEFPRVEL PEPAGDMIST TEKVYVNENE PDVIFVGRI LGPRGITAKO	
Artvanese (*9056 * 10.1.1.5m 0) Artvanese (*9056 * 10.1.1.5m 0) Politike Paties (************************************	a movement full a movement full concernes Options Concernes Options Concernes Options Concernes (pre- liabort rule (concerned) Fraction () Treat trees as rooted? Treat treat trees as rooted? Treat treat trees as rooted? Treat treat tre	Input Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Input View (Converted File) 5 488 C.g., gli-1 MPSCTTPTFO AS-OLEGOS DEPSRESMIT PTSLDDDT-P RKXFPTLETG — DRWP-M C.g., gli-MPSCTTPTFO AS-OLEGOS DEPSRESMIT PTSLDDDT-P RKXFPTLETG — DRWP-M C.g., gli-MPSCTTPTFO VSLDETMA ESPSRS.1 PTSLDEDTSP RKXFPTLETG — DRWP-M C.g., gli-MPSCTTPTFO VSLDETMA ESPSRS.1 PTSLDDETSP RKXFPTLETG — DRWP-M C.g., gli-MPSCTTPTFO VSLDETMA ESPSRS.1 PTSLDCOMSP RKXFPTLETG — DRWP-M C.g., gli-MPSCTTPTFO VSLDETMA ESPSRS.1 PTSLDCOMSP RKXFPTLBT NSLSDRWPPA C.g., gli-MPSCTTPTFO VSLDETMA ESPSRS.1 PTSLDCOMSP RKXFPTDS ISLSPTW-MDSS IPLSORWA- VRDOWSSKR APPPARLSLS TINKOKMSPI SSAYSOTTPKS LLSPAMFNAK SRSFSPTLP PRRDOWSSKR APPPARLSL HT THRVAKASPI SSAYSOTTPKS LLSPAMFNAK SRSFSPTLP PRRDOWSSKR APPPARLSL HT THRVAKASPI SSAYSOTTPKS LLSPAMFNAK SRSFSPTLP PRRDOWSSKR APPPARLSL HT THRVAKASPI SSAYSOTTPKS LLSPAMFNAK SRSFSPTLP PRRDOWSSKR APPPARLSL HTMRVMSPI SSAYSOTTPKS LLSPAMFNAK SRNLESPTLP DEGRVRVAL POTEPRVEL PPPAGNMST TEVYVALKE KANNLTLPH MFNIVERLLD DEIGRVRVAL FOTEPRVEL PEPAGOMIST TEVYVALKE KANNLTLPH MFNIVERLD DEIGRVRVAL FOTEPRVEL PEPAGOMIST TEVYVAND VPDYMFVGRI LGPRGITAKO DEIGRVRVAL FOTEPRVEL PEPAGOMIST TEVYVAND VPDYMFVGRI LGP	
Artvaces (*8055 *10.158m 0) Artvaces (*8055 *10.158m 0) Prot Prot Prot Prot Prot Prot Prove (*8055 *10.158 0) Prove (*8055 *10.158 0) Protee (*8055 *10.158	ten1	Imput Window (PHYLP 3.68 FORMAT-INTERLEAVED SEQUENCES ONLY) Bisot Converted File  4.48  C. gld-1 MPSCTTPTYG VSTOLEBOSS ESPSRSSVMT PTSLDODI-P ROXEPTLETG — DRWP-M CDL_016.0H MPSCTTPTFG A3-OLEIOSO DSPSRESVMT PTSLDODI-P ROXEPTLETG — DRWP-M CDL_04.0H MPSCTTPTFG VSLDETONA ESPSRSSLT PTSLDOETSP RKRFPHDNVPADRWP-S CDL_04.0H MPSCTTPTFG VSLDETONA ESPSRSSLT PTSLDOETSP RKRFPHDSS ILSSRWPPA CDn_016.MPSCTTPTFG VSLDETONA ESPSRSSLT PTSLDOETSP RKSFDIES ILSSRWPPA CDn_016.MPSCR APPPARLSL IT THHORMMSPI SSAYSOTTPIS LLSPANFNEK SRSFSPTLP PRRDOWSSR APPPARLSL HT NHRINMSPI SSAYSOTTPIS LLSPANFNEK SRSFSPTLP PRRDOWSSR APPPARLSL HT NHRIMMSPI SSAYSOTTPIS LLSPANFNEK SRSFSULPRELLD STPLISYGKSS LDSSKSFP PTSP TTEPVEVEX TVEYLADUK EKKHLTLPH MFINNERLD DEIGRWRVAL DOKSENDES HEAVYPANKE PHOYNEVGRI LGPRGNTAKO DEIGRWRVAL FOTFFPRVEL PPAQEMSI TEKVYPANKE PPOYNEVGRI LGPRGNTAKO DEIGRWRVAL FOTFFPRVEL PEPAGOMSI TEKVYPANKE PPOYNEVGRI L	

**Figure 2 JCoDA interface. A)** The JCoDA interface accepts cDNA in FASTA format and JCoDA manages the generation of codon delimited alignments, pairwise or site-based dN/dS calculation through PAML, and graphing of either output (top). The codon delimited alignment tab is shown. Common options are displayed on the main interface and the file menu contains export options. Outputs, such as raw data, alignments, and graphs are displayed in tabbed panes for easy navigation. Graphs for multiple models are retained in the tabbed interface for comparison. For pairwise dN/ dS the user the selects sequences from a list using add/remove buttons. The advanced options check box provides direct access to the codeml control file as a tabbed pane. For site-based analysis JCoDA defaults to implement models M7 (null, neutral) -vs- M8 (alternate, selection) in codeml. The user can set the path for their own tree file or use the PHYLIP graphical interface (PGI) to generate phylogenetic a tree (bottom).



Figure 3 Sample of JCoDA output using sliding window analysis of pairwise *gld-1* dN/dS. Puritying selection dominates the RNA binding GSG domain (amino acids 141-328, *Ce\_gld-1*). All pairwise comparisons suggested relaxed puryfing selection at the N (left) and C-terminal (right) ends. Black bar = GSG domain amino acids 141-328 relative to *Ce\_gld-1*. *Ce = C. elegans, Cbr = C. briggsae, Cre = C. remanei, Cbn = C. brenneri,* and *Cja = C. japonica*. All pairwise comparisons were performed using a window 100 and jump of 10. The graph was saved directly from JCoDA and GSG and bar were added in Microsoft Word.

## Evolutionary selection in the nematode sex determination pathway

The extreme divergence in sex determination pathway components has made their analysis problematic using comparison of distantly related species (e.g. *C. elegans -* vs- *D. melanogaster*) The recent availability of sequence information from nematodes related to *Caenorhabditis elegans* (*C. elegans*) has provided the raw material for the

analysis of sex determination pathway components where sequence divergence has not erased all evidence of directional selection. We used the JCoDA interface to perform a screen for directional selection using all known sex determination pathway components in nematodes closely related to *C. elegans*. We were able to identify genes that have been previously shown (or suspected) to be under positive selection. Furthermore, we identified differences



JCoDA. The dashed box was added in Microsoft Word.

in selection between genes based on their function in the sex determination cascade (Table 1).

Relative to the genome wide *C. elegans/C. briggsae* dN/dS ratio of 0.06 [25], the majority of sex determination genes show elevated levels of non-synonymous substitutions in pairwise comparisons (Table 1, dN/dS, 0.06 -vs-0.23). Considerable variability within each category was present but there are no significant differences in average

pairwise dN/dS scores between the categories. Interestingly, site-based analysis reveals evidence for positive selection in five of nine genes involved exclusively in germ line sex determination and five of nine genes involved in the specification of somatic sex (dosage compensation). In contrast, positive selection was only detected in one gene of seven genes involved in both germ line and somatic pathways (Table 1, grey). Evidence

Gene	Pathway	dN/dS	M7-vs-M8	Proportion of sites with dN/dS>1
fem-1	В	0.13	0.86	0.7
fem-2	В	0.12	3.49	2.6
fem-3	В	0.27	0.99	3.8
her-1	В	0.19	0.09	0.6
laf-1	В	0.04	2.34	1.1
tra-1	В	0.25	7.93	6.0
tra-2	В	0.29	1.60	0.0
fox-1	S	0.06	0.02	0.7
sdc-1	S	0.24	0.72	0.0
sdc-2	S	0.35	3.30	0.0
sdc-3	S	0.45	7.21	7.8
sea-1	S	0.57	11.44	52.5
sea-2	S	0.32	20.11	21.7
sel-10	S	0.06	8.21	1.7
sex-1	S	0.22	12.47	1.6
xol-1	S	0.32	2.66	0.0
fbf-1/2*	G	0.15	3.93	1.3
fog-1	G	0.17	12.30	4.5
fog-2*	G	0.51	48.23	19.2
fog-3	G	0.18	0.67	2.0
gld-1	G	0.07	7.17	1.6
gld-3	G	0.35	36.55	10.9
mog-1	G	0.02	4.87	0.4
mog-6	G	0.04	1.09	0.2
nos-3	G	0.39	50.41	22.5

#### Table 1: Analysis of sex determination pathway genes

Genes are sorted based on functions in germ line and somatic sex determination (B), somatic sex determination (dosage compensation) (S), or germ line sex determination (G). All sex determination gene orthologs used were identified by reciprocal best-hit or conserved synteny (Wormbase). Only genes with large indels and obvious mispredictions that could be resolved were eliminated. dN/dS = average pairwise score for the gene family (Nei & Gojobori, 1986). Models compared for likelihood ratio test (LRT) were M7(dN/dS > 1 disallowed, null) -vs-M8(dN/dS > 1 allowed, alternate), where bold indicates number higher than critical value (Df = 2,  $\chi^2_1 = 5.9915$ , p = 0.05) and rejection of the null. Portion of sites with dN/dS > 1 was calculated as a ratio of selected sites with a probability higher than 0.5 of dN/dS > 1 to total sites. Tree files were based on the ClustalW alignments using the neighbor-joining method (Phylip). See associated FASTA files for ID numbers and species used. (\*) = species specific gene expansion or contains species specific duplications.

of fewer genes that function in multiple pathways with signatures of positive selection likely reflects resistance to change based on additional function constraints.

Positive selection can be difficult to detect when high levels of sequence divergence are present and dS is essentially saturated [26]. Haag and Ackerman (2005) [27] measured nucleotide diversity using sliding windows to detect patches of diversifying selection in *C. remanei fem-3* (**fem**inization in XX and XO). We processed the same dataset with JCoDA and confirmed the presence of positive selection clustered between amino acids 339 and 408 (relative to AY142113), a highly polymorphic region of the protein (Table 2). Also consistent with their data, the *tra-2* region used did not show evidence of positive selection, although the dN/dS ratio was elevated (Table 2, dN/dS, 0.37). As additional intraspecies sequence information is collected, it is likely that the capability to detect positive selection will be significantly enhanced.

Examples of the graphical output generated by JCoDA are shown in Figures 3 and 4. The GLD-1 (defective in

Gene	Pathway	dN/dS	M7-vs-M8	Proportion of sites with dN/dS>1
fem-3	В	0.23	10.42	2.3
tra-2	В	0.37	1.23	0.7

#### Table 2: Analysis of C. remanei fem-3 and tra-2 genes

The *fem-3* and *tra-2* intraspecies data from Haag and Ackerman (2005) were analyzed using both pairwise and site-based methods. Using M7 -vs- M8 models LRT rejects the null for *fem-3* (Df = 2,  $\chi^2_1$  = 9.2103, P < 0.01, bold) but not for *tra-2*. B = involved in somatic and germ line sex determination. dN/dS = average pairwise. Proportion of sites with dN/dS>1 = ratio of selected sites with a probability higher than 0.5 of dN/dS>1 to total sites.

germ line development) RNA binding protein governs the translation of numerous mRNA targets including functioning with FOG-2 in the promotion of spermatogenesis in the hermaphrodite [26,28,29]. As expected, pairwise sliding window analysis reveals that extensive purifying selection dominates regions with homology to the RNA binding GSG (GRP33/Sam68/GLD-1) domain (Figure 3, GSG domain). Interestingly, elevated levels of dN/dS were detected at both the N and C-terminal ends relative to the average pairwise dN/dS for gld-1 (0.07). Using sitebased analysis, we confirmed the presence of residues under positive selection (Figure 4B). Curiously, even though GLD-1 orthologs in C. elegans and C. briggsae share significant amino acid homology (80% identity, 90% similarity) and have a dN/dS ratio consistent with purifying selection, they have opposite functions in sex determination [30]. We speculate that species-specific functions could be explained by at least two mechanisms. First, the extensive conservation in the KH-domain between C. elegans and C. briggsae suggests that at least some of the species differences in GLD-1 function can be explained by a change in mRNA targets. Second, based on the elevated non-synonymous substations at the N and C-terminal ends, we can infer that some changes in GLD-1 function could result from species-specific interactions or regulation.

#### Conclusions

An important part of the analysis of homologous coding sequences is the characterization of evolutionary selection by comparing the rates of synonymous and non-synonymous substitutions. The primary issues with these types of analyses are the difficulty in generating codondelimited alignments, shuttling between programs, and the complexity in configuring programs that are designed to detect positive selection. JCoDA provides a simple platform that integrates several common functions associated with evolutionary analysis of coding sequences and the detection of positive/negative selection. JCoDA is a modular tool built using the BioJava framework, ClustalW, Phylip, and PAML that allows for the rapid assessment and visualization of the pairwise and site-based selection pressure on coding sequences. Using JCoDA we were able to identify multiple sex determination pathway genes with evidence of positive selection based on functional constraints. The JCoDA executable, source code, and tutorial are freely available at <u>http://www.tcnj.edu/</u>~nayaklab/jcoda (Additional File 6).

#### **Availability and requirements**

Project name: JCoDA: A Tool for Detecting Evolutionary Selection

Project home page: <u>http://www.tcnj.edu/~nayaklab/</u>jcoda

Operating system(s): Windows

Programming language: Java

Other requirements: Java Runtime Environment 6.0 (JRE 6) or Java Developer Kit 6 (JDK 6, includes JRE)

License: GPL GNU version 3 for JCoDA or PGI. Please do not violate the copyright or terms of use for ClustalW, Phylip, PAML, and JFreeChart.

Any restrictions to use by non-academics: JCoDA and PGI are provided free for academic use only. Please be aware of the copyright or terms of use for ClustalW, Phylip, PAML, and JFreeChart.

#### Additional material

Additional file 1 Zipped archive that contains the JCoDA/PGI tutorial (tutorial.pdf), readme files (JCoDA readme.txt and PGI readme.txt), and video guides (JCoDA videos guide.txt and Common problems video guide.txt).

Additional file 2 Zipped archive that contains the JCoDA/PGI installation and configuration video (FAQ - Installing and configuring -long-.swf). Not required for all installations.

Additional file 3 Zipped archive that contains videos of checking Java version (FAQ Checking your version of Java -audio-.swf) and a sample of sliding window analysis.

Additional file 4 Zipped archive that contains videos documenting common problems with JCoDA and PGI.

Additional file 5 Zipped archive that contains videos documenting partial functionally with Mac OS X 10.6.

Additional file 6 Zipped archive that contains all JCoDA source code and executable jar files.

#### Authors' contributions

RD: Designed and implemented the Phylip graphical interface, JH: Contributed to the initial development of JCoDA, CL: Designed the JCoDA interface, generated output formats, and implemented integration with ClustalW, SS: Integrated PAML functionality, implemented graphing, and performed sex determination gene analysis, SN: Initiated development of JCoDA, contributed to the analysis of sex determination genes, and coordinated the project. All authors contributed the writing of the manuscript and have approved its final version.

#### Acknowledgements

JH and SN were funded in part by The College of New Jersey Mentored Undergraduate Summer Experience (TCNJ MUSE). We would like to thank Dr. Peter DePasquale (The College of New Jersey, Department of Computer Science) for helpful comments.

#### **Author Details**

<sup>1</sup>Department of Biology, The College of New Jersey, 2000 Pennington Road, Ewing, NJ 08628, USA and <sup>2</sup>Weill Cornell Graduate School of Medical Sciences, 1300 York Ave, Box 65, New York, NY 10065, USA

#### Received: 12 January 2010 Accepted: 27 May 2010 Published: 27 May 2010

#### References

- Pirovano W, Heringa J: Multiple sequence alignment. Methods in molecular biology (Clifton, NJ) 2008, 452:143-161.
- 2. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 1994, **22(22)**:4673-4680.
- Notredame C, Higgins DG, Heringa J: T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* 2000, 302(1):205-217.
- Egan A, Mahurkar A, Crabtree J, Badger JH, Carlton JM, Silva JC: IDEA: Interactive Display for Evolutionary Analyses. *BMC bioinformatics* 2008, 9:524.
- Gupta SK, Kececioglu JD, Schaffer AA: Improving the practical space and time efficiency of the shortest-paths approach to sum-of-pairs multiple sequence alignment. J Comput Biol 1995, 2(3):459-472.
- Lipman DJ, Altschul SF, Kececioglu JD: A tool for multiple sequence alignment. Proceedings of the National Academy of Sciences of the United States of America 1989, 86(12):4412-4415.
- Sze SH, Lu Y, Yang Q: A polynomial time solvable formulation of multiple sequence alignment. *J Comput Biol* 2006, **13(2)**:309-319.
   Yang Z: PAML 4: phylogenetic analysis by maximum likelihood.
- Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution 2007, 24(8):1586-1591.
- Liu G, Uddin M, Islam M, Goodman M, Grossman LI, Romero R, Wildman DE: OCPAT: an online codon-preserved alignment tool for evolutionary genomic analysis of protein coding sequences. *Source code for biology* and medicine 2007, 2:5.
- Endo T, Ikeo K, Gojobori T: Large-scale search for genes on which positive selection may operate. *Molecular biology and evolution* 1996, 13(5):685-690.
- Korber B: Computational Analysis of HIV Molecular Sequences. In HIV Signature and Sequence Variation Analysis Edited by: Learn AGRaGH. Dordrecht, Netherlands: Kluwer Academic Publishers; 2000:55-72.
- Nei M, Gojobori T: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution* 1986, 3(5):418-426.
- Liang H, Zhou W, Landweber LF: SWAKK: a web server for detecting positive selection in proteins using a sliding window substitution rate analysis. Nucleic acids research 2006:W382-384.
- Rozas J, Rozas R: DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* (Oxford, England) 1999, 15(2):174-175.
- Suyama M, Torrents D, Bork P: PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research* 2006;W609-612.

- Holland RC, Down TA, Pocock M, Prlic A, Huen D, James K, Foisy S, Drager A, Yates A, Heuer M, et al.: BioJava: an open-source framework for bioinformatics. Bioinformatics (Oxford, England) 2008, 24(18):2096-2097.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al.: Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England) 2007, 23(21):2947-2948.
- Felsenstein J: PHYLIP Phylogeny Inference Package (Version 3.2). Cladistics 1989, 5:164-166.
- Schmid K, Yang Z: The trouble with sliding windows and the selective pressure in BRCA1. *PLoS ONE* 2008, 3(11):e3746.
- Massingham T, Goldman N: Detecting amino acid sites under positive selection and purifying selection. *Genetics* 2005, 169(3):1753-1762.
- Sawyer SL, Wu LJ, Emerman M, Malik HS: Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proceedings of the National Academy of Sciences of the United States of America 2005, 102(8):2832-2837.
- Zanotto PM, Kallas EG, de Souza RF, Holmes EC: Genealogical evidence for positive selection in the nef gene of HIV-1. *Genetics* 1999, 153(3):1077-1089.
- Yang Z: Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular biology and* evolution 1998, 15(5):568-573.
- 24. Pan C, Kim J, Chen L, Wang Q, Lee C: The HIV positive selection mutation database. *Nucleic acids research* 2007:D371-375.
- Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al.: The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS biology 2003, 1(2):E45.
- Francis R, Maine E, Schedl T: Analysis of the multiple roles of gld-1 in germline development: interactions with the sex determination cascade and the glp-1 signaling pathway. *Genetics* 1995, 139(2):607-630.
- 27. Haag ES, Ackerman AD: Intraspecific variation in fem-3 and tra-2, two rapidly coevolving nematode sex-determining genes. *Gene* 2005, 349:35-42.
- Clifford R, Lee MH, Nayak S, Ohmachi M, Giorgini F, Schedl T: FOG-2, a novel F-box containing protein, associates with the GLD-1 RNA binding protein and directs male sex determination in the C. elegans hermaphrodite germline. *Development (Cambridge, England)* 2000, 127(24):5265-5276.
- Lee MH, Schedl T: Identification of in vivo mRNA targets of GLD-1, a maxi-KH motif containing protein required for C. elegans germ cell development. *Genes & development* 2001, 15(18):2408-2420.
- Nayak S, Goree J, Schedl T: fog-2 and the evolution of self-fertile hermaphroditism in Caenorhabditis. PLoS biology 2005, 3(1):e6.

#### doi: 10.1186/1471-2105-11-284

Cite this article as: Steinway *et al.*, JCoDA: a tool for detecting evolutionary selection *BMC Bioinformatics* 2010, **11**:284

### Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit

