#### Hands on Use Clustalx and MEGA on Windows/MAC machines for sequence analysis

#### Part 1:

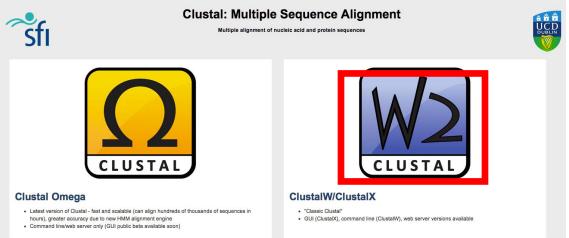
Clustal family tools include clustalw, a command line interface, clustalx, a graphical interface and clustal omega, a recent addition. Clustal was one of the earliest multiple alignment tool, published in 1988.

Higgins, D. G.; Sharp, P. M. (1988). "CLUSTAL: A package for performing multiple sequence alignment on a microcomputer". *Gene* **73** (1): 237–244. <u>doi:10.1016/0378-1119(88)90330-7.PMID</u> <u>3243435</u>

Google	clustal - Q	Set to defai
Scholar	About 85,700 results (0.06 sec)	My Citations
Articles Legal documents	Clustal W and Clustal X version 2.0 MA Larkin, G Blackshields, NP Brown, R Chenna, 2007 - Oxford Univ Press Summary: The Clustal W and Clustal X multiple sequence alignment programs have been completely rewritten in C++. This will facilitate the further development of the alignment algorithms in the future and has allowed proper porting of the programs to the latest	[HTML] from oxfordjo Google Scholar
Any time Since 2013 Since 2012 Since 2009 Custom range	Cited by 6563 Related articles BL Direct All 40 versions Cite  CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix JD Thompson, <u>DG Higgins</u> , TJ Gibson - Nucleic acids research, 1994 - Oxford Univ Press Abstract The sensitivity of the commonly used progressive multiple sequence alignment	[PDF] from nih.gov
Sort by relevance Sort by date	method has been greatly improved for the alignment of divergent protein sequences. Firstly, individual weights are assigned to each sequence in a partial alignment in order to Cited by 42191 Related articles BL Direct All 122 versions Cite More ▼	

Although clustalw is not a favored tool for multiple sequence alignment, clustalx provides a nice graphical interface for visualizing sequence alignment. Here is how to use it:

1. download clustalx from <a href="http://www.clustal.org/">http://www.clustal.org/</a>



2. click on the right side to get to the download page

		<b>N / Clustal)</b> cleic acid and protein sequ	-		
Home Webservers	Download	Documentation	Contact	News	
Webservers	Download	d ClustalW/X			
You don't necessarily have to go through the hassle to install Clustal on your computer. Instead, you can run Clustal online on several servers on the web: • <u>EBI web server</u> • <u>Swiss Institute of Bioinformatics</u>	ClustalX. Precompile the most recent vers <u>here</u> . You can also <u>i</u> The current version	wo flavors: the command-line ed executables for Linux, Ma sion (currently 2.1) along with <u>browse for older versions</u> (C of Clustal 2 is also mirrored ed under the <u>GNU Lesser Gf</u>	ic OS X and Win h the source cc lustalW 1.81, C on the <u>EBI ftp site</u>	e are <u>available for</u>	Viete) of

## 3. connect to the download page

← → C 🗋 www.clustal.org/download/current/

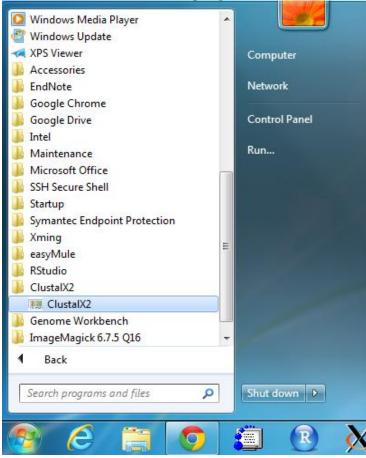
## Index of /download/current

Name	Last modified	Size Description
Parent Directory		-
CHANGELOG	17-Nov-2010 11:59	9.0K
COPYING	17-Nov-2010 11:59	34K
COPYING.LESSER	17-Nov-2010 11:59	7.5K
Readme	17-Nov-2010 11:59	2.0K
Lustalw-2.1-linux-x86_64-libcppstatic.tar.gz	17-Nov-2010 11:59	2.4M
clustalw-2.1-macosx.dmg	17-Nov-2010 11:59	6.5M
clustalw-2.1-win.msi	17-Nov-2010 11:59	1.9M
La <u>clustalw-2.1.tar.gz</u>	10-Dec-2010 07:40	343K
Lustalx-2.1-linux-i686-libcppstatic.tar.gz	17-Nov-2010 11:59	4.7M
clustalx-2.1-macosx.dmg	17-Nov-2010 11:59	12M
clustalx-2.1-win.msi	13-Jan-2011 09:36	4.7M
La clustalx-2.1.tar.gz	10-Dec-2010 07:40	334K

4. install on Windows

B ClustalX2 Setup	
Ð	Welcome to ClustalX2 Setup Wizard
	The Setup Wizard will install ClustalX2 on your computer. Click Next to continue or Cancel to exit the Setup Wizard.
	< Back Next > Cancel

#### 5. find clustalx in the start->program->clustax

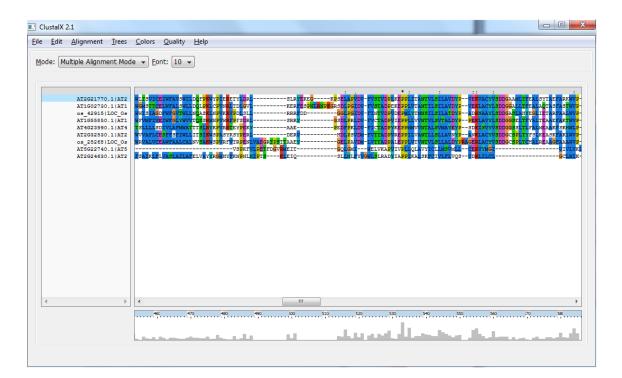


6. this is what clustalx look like

ĺ		alX 2.1		
	<u>E</u> ile	e dit <u>A</u> lignment <u>T</u> rees	<u>C</u> olors <u>Q</u> uality <u>H</u> elp	
	M	ode: Multiple Alignment Mod	de → <u>F</u> ont: 10 →	
			<	4

## 7. load multiple **aligned** sequences

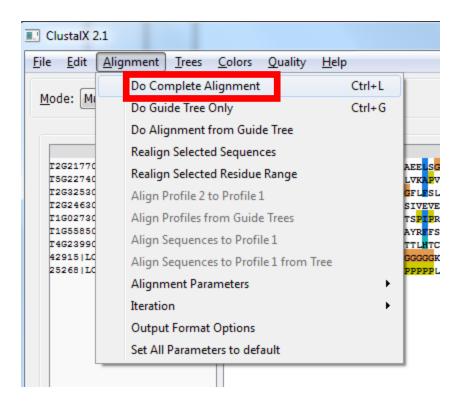
Local >	Microsoft  Windows  Temporar	y Internet Files  Content.IE5  E	GD1235V	✓ <sup>4</sup> → Searc	h EGDT235V	
Irganize 🔻						
🗧 Favorites	Name	Date modified	Туре	Size		
📃 Desktop	adlistrequest[1]	1/31/2013 11:20 AM	JScript Script File	7 KB		
🗼 Downloads	adlistrequest[2]	1/31/2013 2:19 PM	JScript Script File	8 KB		
🚹 Google Drive	cesa-pr.fa[1].aln	2/24/2013 5:14 PM	ALN File	18 KB		
Recent Places	cesa-pr.fa[2].aln	2/24/2013 5:14 PM	ALN File	18 KB		
	💽 x[1]	2/2/2013 3:04 PM	GIF image	1 KB		
🗃 Libraries	💽 x[2]	2/2/2013 3:04 PM	GIF image	1 KB		
	💽 x[3]	2/2/2013 3:04 PM	GIF image	1 KB		
🖳 Computer						
🏜 Local Disk (C:)						
👝 Docs (E:)						
📢 My Web Sites on MS						
📮 Network						
				✓ All File		



#### 8. you can also load in multiple **unaligned** sequences

ClustalX 2.1					range digeneral i	
Eile <u>E</u> dit <u>A</u> lignment <u>T</u> rees	<u>C</u> olors <u>Q</u> uality <u>H</u> elp					
Mode: Multiple Alignment Mo	de → <u>F</u> ont: 10 →					
		ALVSVSQKV <mark>P</mark> EDGWTMQD <mark>GTPWPG</mark> NNV			GAMNSLIR <mark>VSAVL</mark> SN <mark>AP</mark> YLIN <mark>V</mark> DC	DHY NNSKAIRE MC FMMD
AT5G22740.1 AT5G22740.	I C TAKLOSGQSAKGNIKGIKRFP	RIFKLPDRLNTLELGFAAFLFVCGCYD	FVHGKNNYFIYLFLQTMSFFISGLG	II <mark>GTYVP</mark> S		
AT5G22740.1 AT5G22740. AT2G32530.1 AT2G32530.	I C TAKLESEQSAKENTKEIKREP		FV <mark>HGKNNYFIYL</mark> FLQ <b>TMSFF</b> ISGLG RWATGLLEVLFNKQSPLIGMFCRKIF	I <mark>GTYVP</mark> S FRQSLAYLYIFTW <mark>GL</mark> RSI	PELTYCL PAYCLIHNAAL FPROV	YLGIVYTLVGMHCLYSLWE
AT5G22740.1 AT5G22740. AT2G32530.1 AT2G32530. AT2G24630.1 AT2G24630. AT1G02730.1 AT1G02730.	IIG TAKLGSGQSAKGNTKGIKRFP IIG LYESTAEDANTSIGIHSRGWT IIG KQQHRWHS <mark>GP</mark> MOLFRLC <mark>R</mark> SI IIG AQTASFASTWV <mark>P</mark> FCREHNIEP	RIFRIPORINTIELGFAAFIFVCGCYD 339 I39R <mark>PB</mark> AFIGAMPPGGPEAMLQOR 1138 IAWWKKANLIL FFLIRRIIP RN <mark>BEAYFG</mark> ORRNFIRMKVRIDFVRERR	)FVHGKNNYFIYL FLQTMSFFISGLG RWATGLLEVLENKQS <mark>PLIGM</mark> FCRKIF YSFTLCVILPITMFVPEAEL <mark>P</mark> IWVI RVKREYDEFKVRINSL <mark>PEAI</mark> RRRSD	IIGTYV <mark>P</mark> S FRQSLAYLYIFTWGLRSI CYV <mark>PIF</mark> MSLLNILPAPK <mark>S</mark> YNVHEELRAKKKQMEMMM	PELTYCLE PAYCLEHNAAL FPRGV FPFIVPYLLFENTMSVIKENAMVS GNNPQETVIV <mark>FKAIWM</mark> SD <mark>GSHWPG</mark>	YLGIVVTLVGMHCLYSLWEF Glfqlgsyewivtkkagrs: Twss <mark>get</mark> dnsrgdhagiiqa
AT5G22740.1 AT5G22740. AT2G22530.1 AT2G22530. AT2G24630.1 AT2G24630. AT1G52730.1 AT1G2730. AT1G55850.1 AT1G55850.	IIC TAKLESEQSAKENTKEIKREP IIC LYESTAEDANTSIEHSREWT IIC KQQHRWHSEPMQLFRLCERSI IIC AQTASFASTWVPFCRKHNIEP IIC EWKSAYLNPEKQAFLEVAFTN	RIFL PORLNILLGFAAFLFVCGCYD 35%ISFRPARLGAAPPGGPEAMLOOR 1176KTAMWKKANLIL FLIRKILGF 18NPFAYFGORNNFLENKVELDFVRERR 1110MLWOORRWSEGDFOIMLSKYS <mark>P</mark> VW	)FVHGKNNYFIYLFLQTMSFFISGLG RWHTGLLEVLPNKQSFFIGHFCRKF YSFLLCVLFJTMFVFEAELFIW RWRRFYDFFKVFINSJFAIRRSD YGKGKISLGLILGYCCYCLWAFSS	IGIYV <mark>P</mark> S ERQSLAYLYIFINGLRSI CYV <mark>P</mark> IEMSLLNILPAPKS YNVHEELRAKROMEMOM VLIYSVLTSLCLER <mark>GIPL</mark>	PELTYCLLPAYCLIHNAALFPRGV FPFTVPYLLFENTMSVTKENAMVS GNNPQETVIVPRATAMSDGSHMPG FPKVSSSWFIPFGYVTVAATAYSL	YLGIVVILVGMHCLYSLWER Glfolgisyewivikagrs: Twssgeidnsrodhagiiqa Aefiwcggifrgwwneorwy
AT5622740.1 AT5622740. AT2632530.1 AT2632530. AT2624630.1 AT2624630. AT1602730.1 AT1602730. AT1655850.1 AT1655850. AT4623990.1 AT4623990.	IIC TAKLESGQSARGNTKGIKRFD IIC LYESTAEDANTSIGIHSRGWT IIC RQQHNWHSGFMQLFRLCHRSII IIC AQTASPASTWVPFCRKHNIED IIC GWKSAYINDERGAFLGVATU. IIC SKIGFRYGSLVEDYYTGYRLH	RIFKLPDRINTLELGFAAFIFVGGCYD 33 YIS RPPAFIGMPPGGPFALQOR LISRTAMMKRANIIIFFLRKTIIF RN <mark>EAYFS</mark> ORNFERMKVRLDFVRERR LHGMLQORRMSEGDFOIMISKSSYW CCE <mark>WRSVFCR</mark> KRAAFCGDS <mark>B</mark> KSLIDV	HFVHGKNNYFIYLFLCTMSFFISGLG RWATGLEVLENKQSFLIGHFCKKI YSTLCVILFTMFVFEALLTWI RVRREYDEFKVRINSLFAIRRSSS YSCQKRWALGLEVAISRYS <mark>F</mark> IYG	IGTYVPS FRQSLAYLYIFTWGIRS CYVPIFMSLINILPAPKS YNVHEELRAKKKQMEMOM VLIYSVITSLCIFKGIPL KSMGLVTGVGYCQYACWA	PELEYCLI PAYCLI HNAAL FPKGV FPELEYEYLLE ENTMSVIKENAMVS GNFQETVI PRATAMSGSHWGG FFKVSSSWFI PEGYVVARAYS FMSLPELYYGFL PQLALLAQSSVF	YLGIVVILVGMHCLYSLWEF Glyglgsyewiytkkars: Twssgeidnsrchagiiqa Aefiwcggffrgwwneghwi Pkssdpwewlyivlflgayg
AT5622740.1   AT5622740. AT2632530.1   AT2632530. AT2634630.1   AT26326480. AT1602730.1   AT1602730. AT1655550.1   AT1655550. AT4623990.1   AT4623990. o=_\$2151   DC_0=07g36610.	IIC TARLESS QSARONTKGIRRED IIC LYESTAEDANTS I GIHARGWI IIC RQQHRHASPMQLERLCHS I IIC AQTASFASTWOPCRKHNIED IIC GHKSANLWERQAFLGVATH IIC SKIGSVESSUEDVYTOKTH IIC LELWEVQSSVESTCKRRM	RIFL PORLNILLGFAAFLFVCGCYD 35%ISFRPARLGAAPPGGPEAMLOOR 1176KTAMWKKANLIL FLIRKILGF 18NPFAYFGORNNFLENKVELDFVRERR 1110MLWOORRWSEGDFOIMLSKYS <mark>P</mark> VW	FVHGENNYFIYLFLQIMSFTISGLG IRMATGULEVLFNKQSPLIGHFCRKI YSGILFVLFLTMYVDFALLBIKV RVKREYDFFKVRINSLFFALRRSD YGWGULSLGLUGVGCVCLA <mark>B</mark> SSL VSQQKRMAIGLLEVALSRYSJIYG NSK <mark>HFI</mark> NIFAAASQCRSIS <mark>B</mark> AAS	IGTYVPS PRQSLAYLYIFTNGLRSI CYVPIFMSLINILPAFRS VNVHEELRARKROMEMOM VLIVSVLTSLCLERGIPL KSMGLVTFVGVCQVACMA LDETAAMAEVEEVNTCAY	PELIVCL PAYCI HNAA FPRGV PETVYLLEENTHSVIRENAMVS GNNPQETVIVERIMSDCSHNPG FRVSSKFIFFGVVTVARASI FNSLPIIVGELPQIALLQSSVF EGTERGGVGVVTDIALLQSSVF	YLGIVVILVGMHCLYSLNEF GLEQLGSYYENIVTKAGRS TWSSGEIDNSRCHAGIQA Abffweggfrgmwnogrmu FRSSDFWFNLYIVILGAYG FRLHRRGWRSMYCAMEPDAF

and do clustalw alignment



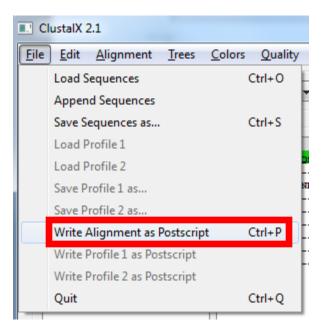
you will be asked where you want to save the alignment results (an alignment and a dendrogram)

Output Guide Tree	2:
Windows/Tempor	ary Internet Files/Content.IE5/TGB3B3D5/cesa-pr[1].dnd Browse
Output Alignment	Files
Clustal:	Iternet Files/Content.IE5/TGB3B3D5/cesa-pr[1].aln Browse
ОК	Cancel

#### finished

ClustalX 2.1	
<u>File Edit A</u> lignment <u>T</u> rees	<u>Colors</u> <u>Q</u> uality <u>H</u> elp
Mode: Multiple Alignment Mode	▼ Eont: 10 ▼
F1G02730.1 AT1G02730.1 cslD 42915 LOC 0s07q36610.1 cslF	
T2G21770.1 AT2G21770.1 cesA T2G32530.1 AT2G32530.1 cs1B	MITGGELIAGSHERVETVLIVADITARIRBARELSONICKICHDETELIDHGEPPITAMECAPPICRECHEVERBEGHDARPOG
25268 LOC_0=04g35020.1 c=1H F1G55850.1 AT1G55850.1 c=1E	
T4G23990.1 AT4G23990.1 cs1G T5G22740.1 AT5G22740.1 cs1A	
T2G24630.1 AT2G24630.1 cslC	

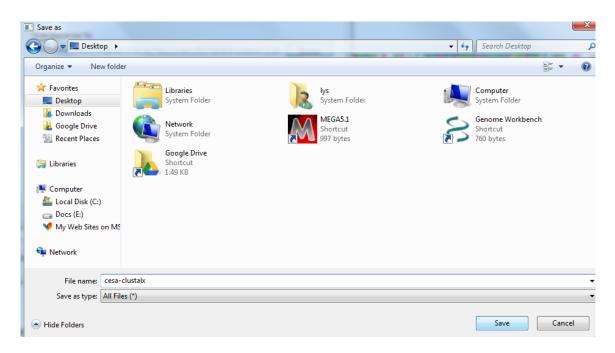
9. you can save the alignment as a postscript format file



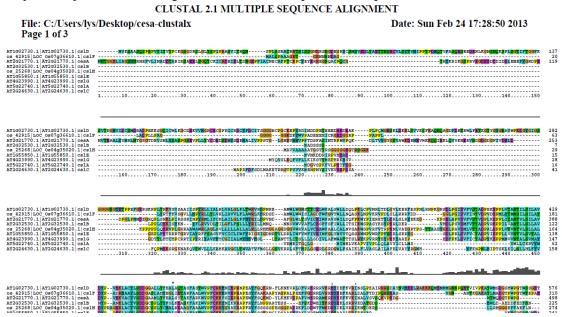
#### you may choose different options

] clustalx				? ×
Write Sequences To	:			
/Microsoft/Windows	s/Temporary Internet Files/C	ontent.IE5/TGB3B3D5/ces	a-pr[1].ps	Browse
PS Colors File:				
Page Size	A4 🗸	Orientation	Landscape	•
Print Header:	Yes 👻	Fint Quality Curve:	Yes	•
Print Ruler:	Yes 👻	Fint Residue Numbers:	Yes	•
Resize to fit page:	Yes 👻	]		
Print from position	1	t	1249	
Use block length:	150			
ок	Cancel			

I am gonna save it in the desktop folder



#### 10. you may use **acrobat pro** to convert the ps file to a pdf file



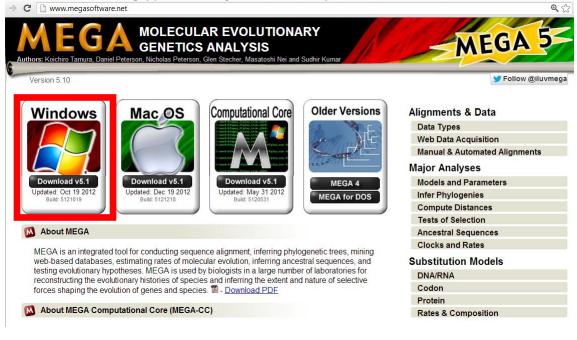
#### **Part 2:** MEGA5: Molecular Evolutionary Genetics Analysis version 5

MEGA is an integrated tool for conducting sequence alignment, inferring phylogenetic trees, mining webbased databases, estimating rates of molecular evolution, inferring ancestral sequences, and testing evolutionary hypotheses. MEGA is used by biologists in a large number of laboratories for reconstructing the evolutionary histories of species and inferring the extent and nature of selective forces shaping the evolution of genes and species

Mega was developed as a software with GUI, but recently it released a command line version to facilitate large scale analyses using terminals: http://www.kumarlab.net/pdf\_new/KumarTamura12a.pdf

Scholar	About 20,400 results (0.11 sec)	Ny Citations
Articles Legal documents	MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0 <u>K Tamura</u> . JDudley, M Nel, <u>S Kumar</u> - Molecular biology and evolution, 2007 - SMBE Abstract We announce the release of the fourth version of MEGA software, which expands on the existing facilities for editing DNA sequence data from autosequencers, mining Web-	[нтмь] from 9med.net Google Scholar
Any time Since 2013 Since 2012 Since 2009 Custom range	databases, performing automatic and manual sequence alignment, analyzing sequence Cited by 17875 Related articles BL Direct All 15 versions Cite MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods <u>K Tamura</u> , D Peterson, N Peterson, G Stecher Molecular biology and, 2011 - SMBE	[HTML] from oxfordjourr Google Scholar
Sort by relevance Sort by date	Here, we announce the release of Molecular Evolutionary Genetics Analysis version 5 (MEGA5), which is a user ancestral states and sequences (along with probabilities), and estimating evolutionary rates site-by This version of MEGA is intended for the Windows platform, and it Cited by 4162 Related articles All 14 versions Cite	
<ul> <li>✓ include patents</li> <li>✓ include citations</li> </ul>	MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences <u>S Kumar.</u> M Nei, <u>J Dudley. K Tamura</u> - Briefings in bioinformatics, 2008 - Oxford Univ Press It was made available over the Internet (http://www.megasoftware.net) and was downloaded by	[HTML] from oxfordjourr Google Scholar
Create alert	In MEGA 4, we expanded the transparency of choices and assumptions by adding a new Caption MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 Cited by 1431 Related articles BL Direct All 27 versions Cite More →	

#### 1. find MEGA at http://www.megasoftware.net/



2. it's free, but you need to fill out an on-line form to download

## I also put a recently downloaded file at <u>http://cys.bios.niu.edu/yyin/teach/MEGA5.10\_Setup.exe</u>

Version 5.10 Version 5.10 IEGA 5.10 (Windows	erson, Glen Stecher, Masatoshi Nei and Sudhir Kumar
	.)
IEGA 5.10 (Windows	
	5)
IEGA 5 is provided FREE for use in research	n and education (see terms below).
To download MEGA 5 , please fill in the follow will allow us to assess the impact of this softw	wing information. This will enable us to provide information about future upgr ware.
mportant Notice: By downloading MEGA 5, y	you are agreeing not to redistribute the MEGA 5 software and associated ma
his version is for Windows only. It is NOT inter	nded to be used on Mac or LINUX systems, as it has not been optimized for
	Last Name*
	First Name*
	E-mail Address*
Please make sure that you provide valid in Ve will send you a download link by the e-	

#### 3. follow the instructions to install

🔂 Setup - MEGA5.1	
Typ Setup - MEGAS.I	Welcome to the MEGA5.1 Setup Wizard This will install MEGA5.1 on your computer. It is recommended that you close all other applications before continuing. Click Next to continue, or Cancel to exit Setup.
GCCGTATC	Next > Cancel

4. this is what MEGA looks like

	MEGA 5.10 (5121019)			• X
ĺ				
	😂 Open A File/Session	אַ <u>ת</u> ּיּי ר <u>ק</u> ייי <mark>הַיָּיָי</mark> Distance ▼ Diversity ▼ Phylogeny ▼ User Tree ▼	Ancestors Selection Rates Clocks	٠.
I		Distance Diversity Phylogeny User Tree	Ancestors Selection Rates Clocks	Links
	🚇 Edit a Text File			
l	Mes Convert File Format to MEGA			
	Printer Setup			
	Exit MEGA Alt+X			
	First time User? Tutorial Exam		, Customize ToolBar 👻	
	MEGA release #5121019			:

### 5. let's load an **aligned** sequence file

M Open a File	manufe mathematics the		×
Look in:	Documents 🗨	+ 🗈 📸 🖬 -	
Ca.	Name	Date modified	Туре
Recent Places	\mu easyMule1.0	2/8/2013 12:43 PM	File folder
	b EndNote	1/28/2013 8:13 PM	File folder
	📕 EndNote	1/15/2013 11:55 AM	File folder
Desktop	i R	2/11/2013 4:24 PM	File folder
<u> </u>	퉬 Thunder Network	2/8/2013 12:40 PM	File folder
6 <b>3</b> 3.	Rhistory	2/24/2013 2:05 PM	RHISTOR
Libraries			
Computer			
Network	•		•
	File name: ;ys.bios.niu.edu/yyin/teach/PBB/ce	sa-pr.fa.aln 💌	Open
	Files of type:	•	Cancel
			Help

6. a new window called sequence editor will appear

M5: Text File Editor and Format Converter	
File Edit Search Display Utilities	
cesa-pr.fa[1].ain	۲
CLUSTAL format alignment by MAFFT L-INS-1 (v6.850b)	
AT2G21770.1 AT2 M	
AT2G21770.1 AT2 N AT1G02730.1 AT1 NRRSNGDEGRYCSMSVEDLTAETTNSECVLSYTVHIPPTPDHQTVFASQESEEDEMLKGN os_42915 LOC_Os AT1G55850.1 AT1 AT4G23990.1 AT4	-
1:1 modified Insert	1

## 7. convert the aln format to meg format

M5: Select File and	Format	×
Data file to convert:		
C:\Users\lys\AppD	ata\Local\Microsoft\Windows\TemporaryIr	nternet File 😅
Data Format: .aln	(CLUSTAL)	•
	OK X Cancel	<u>? H</u> elp

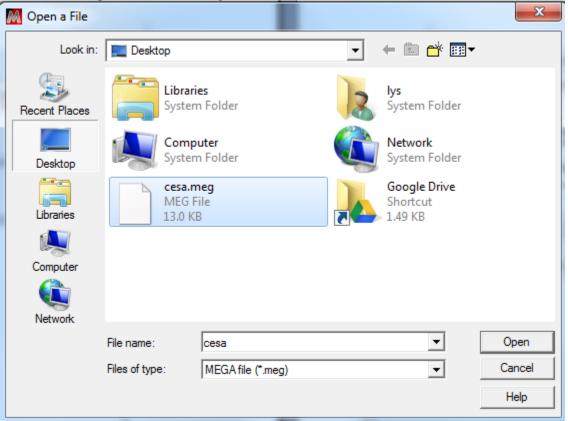
8. save it as a meg file

M Save As	_	and such	tile for		×
Save in:	Desktop		•	⇐ 🛍 📸 ◄	
Recent Places	Libra Syste	a <b>ries</b> em Folder	13	<b>lys</b> System Folder	
Desktop		n <b>puter</b> em Folder		<b>Network</b> System Folder	
Libraries		<b>gle Drive</b> tcut KB			
i 🌉					
Computer					
Network					
	File name:	cesa.meg		•	Save
	Save as type:			•	Cancel
MEGA 5.1				-	<b>—</b> ×
File conversion complete. P the screen before proceedin		y not be in the exact format th	at MEGA expects for importing,	Therefore, you must check the c	onverted results carefully on

9. what meg format look like:

File Edit Search Display Utilities	
<b>28</b> 20 10 10 10 10 10 10 10 10 10 10 10 10 10	
cesa-pr.fa[1].aln cesa.meg	2
#Mega	
!Title cesa-pr.fa[1].aln;	
#AT2G21770.1 AT2	
MNTGGRLIAGSHNRNEFVLI	
N	
KICRDEIELTDNGEPF	
IACNECAFPTCRPCYEYERREGNQACPQCGTRYKRIKGSPRVEGDEEDDDIDDLEHEFYG	
MDPEHVTEAALYYMRLNTGRGTDEVSHLYSASPGSEVPLLTYCDEDSDMYSDRHALIVPP	
STGLGNRVHHVPFTDSFASIHTRPMVPQKDLTVYGYGSVAW-KDRMEVWKKQQ	
IEKLQVVKNERVNDGDGDGFIVDELDDPGLPMMDEG-RQPLSRKLPIRSSRINPY	
RMLIFCRLAILGLFFHYRILHPVNDAFGLWLTSVICEIWFAVSWILDQFPKWYPIE	
RETYLDRLSLRYEKEGKPSELAPVDV-FVSTVDPLKEPPLITANT	
VLSILAVDYPVEKVACYVSDDGAAMLTFEALSYTAEFARKWVP-FCKKFSIEPRAPEW	
YFSQKMD-YLKHKVDPAFVMERRAMKRDYEEFKVKINALVSVSQK	
NNVRDHPGMIQVFLGHSGVC	
DMDGNELPRLVYVSREKRPGFDHHKKAGAMNSLIRVSAVLSNAPY	
LLNVDCDHYINNSKAIREAMCFMMDPQ-SGKKICYVQFPQRFDGIDRHDRYSNRNVVF	
FDINMKGLDGIQGPIYVGTGCVFRRQALYGFDAPKKKQPPGRTCNCWPKWCCLCCGMR	

10. now let's open the saved meg file

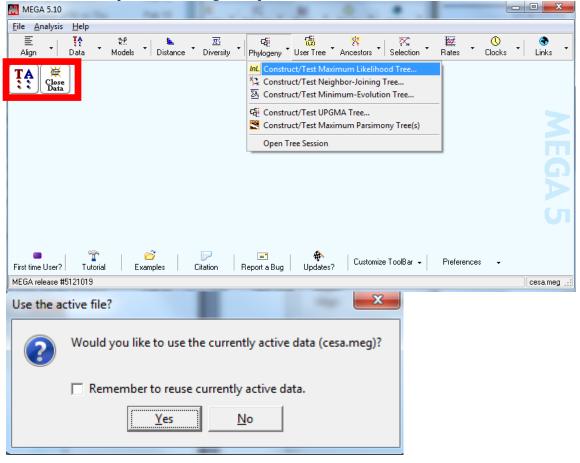


11. you will be asked for the data type

M5: Input Data	×
Data Type	Missing Data 2
Protein Sequences	
	Alignment Gap - Identical Symbol -
	✓ OK ★ Cancel ? <u>H</u> elp

12. the data is loaded; we can build the tree now

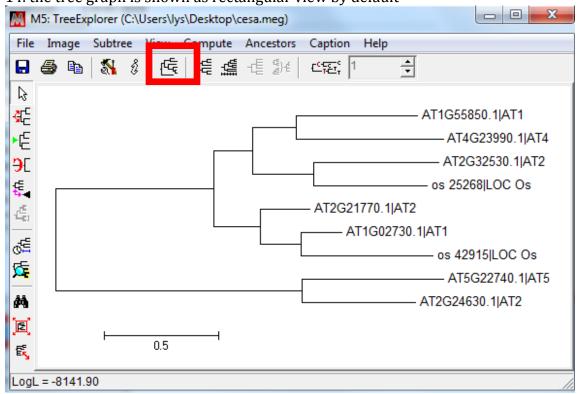
You may choose from a list of different building algorithms Basically, maximum likelihood is the most accurate but also the slowest Neighbor-joining and maximum parsimony are also very popular and faster if you have over 50 sequences or longer sequences



13. you may choose parameters for tree building, but let's try default first (the default ones are the fastest but less accurate)

Options Summary		
Option	Selection	
Analysis	Phylogeny Reconstruction	
Statistical Method	Maximum Likelihood	
Phylogeny Test		
Test of Phylogeny	None	
No. of Bootstrap Replications	Not Applicable	
Substitution Model		
Substitutions Type	Amino acid	
Model/Method	Jones-Taylor-Thornton (JTT) model	
Rates and Patterns		
Rates among Sites	Uniform rates	
No of Discrete Gamma Categories	Not Applicable	
Data Subset to Use		
Gaps/Missing Data Treatment	Complete deletion	
Site Coverage Cutolf (%)	NotApplicable	
Tree Inference Options		
ML Heuristic Method	Nearest-Neighbor-Interchange (NNI)	
Initial Tree for ML	Make initial tree automatically (Default - NJ/BioNJ)	
Initial Tree File	NotApplicable	
Branch Swap Filter	Very Strong	
System Resource Usage		
Number of Threads	1	

M5: Progress	Colors Turney	
Progress	100%	
Details 🔻		Stop
Status / Options		
Run Status		
Start time	2/25/2013 11:02:23 AM	
Status	Optimizing final tree	
Log Likelihood	-8141.9604	
Analysis Options		
Analysis		
Analysis		- Phylogeny Reconstruction
Statistical	Method	- Maximum Likelihood
Phylogeny Test	t	
Test of Phyl	logeny	- None
Substitution N	Model (	
	из Туре	
		- Jones-Taylor-Thornton (JI
Dates and Datt	ame	



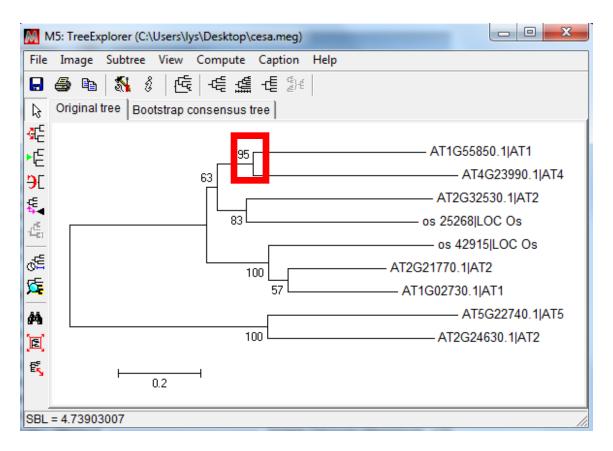
#### 14. the tree graph is shown as rectangular view by default

15. how do we have statistical values on the internal nodes?

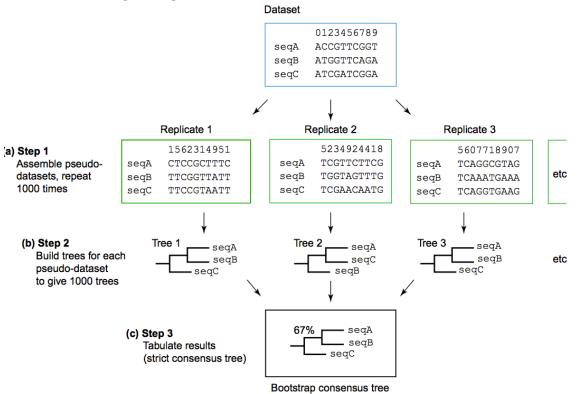
This time we want to go back to choose neighbor-joining algorithm because it is much faster than maximum likelihood. Here we also want to choose bootstrap method to test the phylogeny, then we will have statistical values for each node.

M5: Analysis Preferences	
Options Summary	
Option	Selection
Analysis	Phylogeny Reconstruction
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
Phylogeny Test	
Test of Phylogeny	Bootstrap method
No. of Bootstrap Replications	500
Substitution Model	
Substitutions Type	Amino acid
Model/Method	Poisson model
Rates and Patterns	
Rates among Sites	Uniform rates
Gamma Parameter	Not Applicable
Pattern among Lineages	Same (Homogeneous)
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
Site Coverage Cutoff (%)	Not Applicable
	Compute X Cancel ? Help

16. now we have the bootstrapped neighbor-joining tree.



How are bootstrap tests performed?

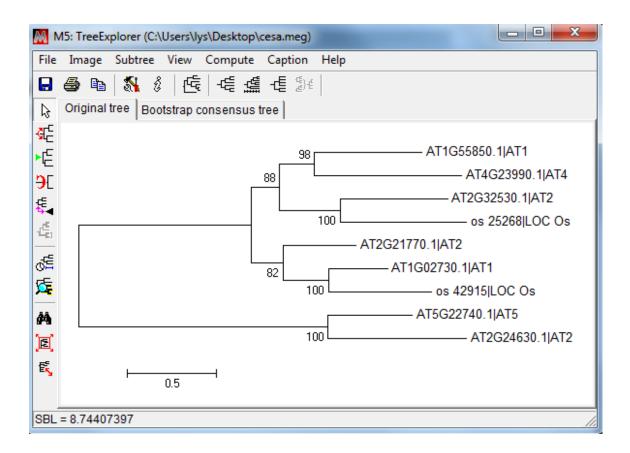


TRENDS in Genetics

M M5: Analysis Preferences	iodels to estimate the tree	
Options Summary		
Option	Selection	
Analysis	Phylogeny Reconstruction	
Scope	All Selected Taxa	
Statistical Method	Neighbor-joining	
Phylogeny Test		
Test of Phylogeny	Bootstrap method	
No. of Bootstrap Replications	500	
Substitution Model		
Substitutions Type	Amino acid	
Model/Method	Jones-Taylor-Thornton (JTT) model	
Rates and Patterns		
Rates among Sites	Gamma Distributed (G)	
Gamma Parameter	4	
Pattern among Lineages	Same (Homogeneous)	
Data Subset to Use		
Gaps/Missing Data Treatment	Pairwise deletion	•
Site Coverage Cutolf (2)	Not Applicable	
	✓ C <u>o</u> mpute	X Cancel ? Help

17. if we want to have more accurate tree, we can tune more parameters to use sophisticated statistical models to estimate the tree

Note the tree topology and bootstrap supports changed?



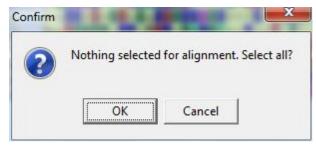
# 18. If we want to start from unaligned sequences and use built-in clustalw or MUSCLE to perform sequence alignment: MEGA 5.10 (5121019)

Dpen A File/Session	pistance - Diversity - Phylogeny - User Tr	ree * Ancestors * Selection * Rates	<u> </u>
open a Recently Oscal life	Distance * Diversity * Phylogeny * User Tr	ree Ancestors Selection Rates	🕇 Clocks 🎽 Links 🕇
🚰 Edit a Text File			
Mes Convert File Format to MEGA			
Printer Setup	-		
	-		
Exit MEGA Alt+X			
		Customize ToolBar 🗸	
	nples   Citation   Report a Bug   Upd	lates?	10.00
MEGA release #5121019			

MEGA 5.10 (5121019)	
<u>Eile A</u> nalysis <u>H</u> elp	
E J TA Models Distance Diversity Phylogeny User Tree Ancestors Selection Rates	© Clocks ▼
Analyze or Align File?	MEGA 5
First time User? Tutorial Examples Citation Report a Bug Updates? Customize ToolBar - Preference	- 26

#### The alignment editor:

M5: Alignment Explorer (cesa-pr[2].fa)	
Data Edit Search Alignment Web Sequencer Display Help	
] 🗅 😅 🖬 🎬   🗮 🌆 🎆   W 💜   💥 🤽 -  ] 🗠 🗈 👗 💼 🗙 🛪 🛪 I 🐿 🙋   🖣 🕨	M 🏘 🏘 🔥
Protein Sequences	
Species/Abbrv +	
1. AT2G21770.1 AT2G21770.1 cesA Ling Griling Hinn Fviin Andra Ing Anglis Cont	I C R D E I E L T D N G E I
2. AT5G22740.1 AT5G22740.1 cslA MOGVORKEVLDI FOGVAMIIGOLGMINIUVKADVIVI	LLQLAV <mark>Y</mark> ICLLMSV
3. AT2G32530.1 AT2G32530.1 cs1B MADSESSLEPLCEKISYKNYFLRVVDLTILGFLFSLLLY	ILLMNQNNSVWVV
4. AT2G24630.1   AT2G24630.1   cslC MAPPF DE LUAKSTRRGTPVVV MENENYSIVSVEDPD	AFQPMEKSRGKNA
5. AT1G02730.1   AT1G02730.1   cs1D MVKSAASQEPEVTIVTECKGGGDBGLGLGEFIFEASVI 6. AT1G55850.1   AT1G55850.1   cs1E MVKKDQAIAPVHSAGGPLFETRREGEVIAYRFFSASVI	FVCICLIWFYRIGE
7. ATIGSSSSSSTIRTIGSSSSSTIRTIGSSSSSTIRTIGEN AV	RTIPYRIYAVFHT
3. os 42915/LOC Os07q36610.1/cs1F MALCPAAAGRIGRNNNNDAGLADPLLPAGGGGGGGGKIKY	
9. os 25268   LOC Os04g35020.1   cs1H MAVVAAAAATGSTTRSGGGGGGGGGTRSGRKKPPPPLQER	
<	4
Site # 1 👻 🕫 with C w/o Gaps	



## Select MUSCLE alignment

M5: MUSCLE - AppLink			
Option	Selection		
Presets	None		
Gap Penalties			
Gap Open	-2.9		
Gap Extend	0		
Hydrophobicity Multiplier	1.2		
Memory/Iterations			
Max Memory in MB	808		
Max Iterations	8		
More Advanced Options			
Clustering Method (Iteration 1,2)	UPGMB		
Clustering Method (Other Iterations)	UPGMB		
Min Diag Length (lambda)	24		
Genetic Code (when using cDNA)	Standard		
Alignment Info	MUSCLE Citation: Edgar, Robert C. (2004), MUSCLE:		
	multiple sequence alignment with high accuracy and		
high throughput, Nucleic Acids Research 32(5), 1792-1797.			
	Compute Cancel ? Help Restore Defaults		

M M5: Alignment Explorer (cesa-pr[2].fa)			x
Data Edit Search Alignment Web Seguencer Display Help			
D 😅 🖬 🛱   🧮 🌆 🎆   W 😏   🎽 🏗     🗠 🗈 🌡 💼 🗙 📯   🍓 🖄   4 🕨	M 4	10 AQ	<b>8</b>
Protein Sequences	12		
Species/Abbrv			
1. AT1G02730.1 AT1G02730.1 cs1D MVKSAASOSPSPVTITVTPCKGSGORSLGLTSPIPRASVI	INQNSP	LSSR	
2. AT1G55850.1   AT1G55850.1   cslE 3. AT2G21770.1   AT2G21770.1   cesA	M		
		INA	
4. AT2G24630.1 AT2G24630.1 cslCMA	<u> </u>		
5. AT2G32530.1  AT2G32530.1   cs1B			
6. AT4G23990.1 AT4G23990.1 cslGMX V I 7. AT5G22740.1 AT5G22740.1 cslA			DGV
8. os 25268 LOC 0s04q35020.1   cs1H	<mark>MAVV</mark>		
9. os 42915/LOC 0s07g36610.1/cs1F		LSPA	
3. 09 42315[100 050/g38610.1]C317	MA		AAB
			•
Site # 1129 🚔 💿 with 🗅 w/o Gaps			: