

Performing ClustalW protein alignments directly from the Orthology and Synteny table within the gene record pages

Summary of the new feature: ClustalW is a tool that aligns three or more sequences. EuPathDB gene record pages now have a feature that allows you to compare the sequence of interest (e.g. *Neurospora crassa* gene NCU00007) to a custom list of orthologs in other species (e.g. orthologs in *Fusarium* species). The main output is visualization of aligned sequences as text followed by a .dnd file at the bottom of the output screen. A .dnd file is useful for building trees via various free web applications. Below is a brief tutorial on how to perform protein alignment for selected species with ClustalW that can be activated directly from the Orthologs table on any gene record page of the EuPathDB databases.

1. Navigate to a gene record page for NCU00007 from www.fungidb.org

NCU00007 pH-response regulator protein palH/rim-21

Name: prr-4
Type: protein coding
Chromosome: III
Location: CM002238:2,256,790..2,260,238(+)
Species: *Neurospora crassa*
Strain: OR74A
Status: Reference Strain

Shortcuts
Synteny BLAT Alignments SNPs Transcriptomics Protein Features
Also see NCU00007 in the [Genome Browser](#) or [Protein Browser](#)

2. Select the **Orthology and Synteny** menu from the Contents menu on the left and select orthologs for ClustalW alignment:

- Use the *search box* at the top to bring up orthologs by species name (e.g. *Fusarium*)
- Select orthologs for ClustalW sequence alignment (e.g. in this example all *Fusarium* orthologs were selected by clicking on the **Select All** button)

Contents
expand all | collapse all
Search section names...
1 Gene models
2 Annotation, curation and identifiers
3 Link outs
4 Genomic Location
5 Literature
6 Taxonomy
● 7 Orthology and synteny
8 Phenotype
9 Genetic variation
10 Transcriptomics
11 Sequence analysis
12 Sequences
13 Structure analysis
14 Protein features and properties
15 Function prediction
16 Pathways and interactions
17 Immunology
expand all | collapse all

Orthologs and Paralogs within EuPathDB Data sets
Showing 9 of 69 rows
Bring up orthologs by species name
Fusarium

Protein clustalW	Gene	Organism	Product	Is syntenic	has comments
<input type="checkbox"/>	FFUJ_08659	<i>Fusarium fujikuroi</i> IMI 58289	related to pH signal transduction protein PalH	yes	no
<input type="checkbox"/>	FGRAAPH1_01G18125	<i>Fusarium graminearum</i> PH-1	pH-response regulator protein palH/rim-21	yes	no
<input type="checkbox"/>	FOCI_g10008702	<i>Fusarium oxysporum</i> f. sp. cubense race 1	pH-response regulator protein palH/rim-21	yes	no
<input type="checkbox"/>	FOCA_g10008637	<i>Fusarium oxysporum</i> f. sp. cubense race 4	pH-response regulator protein palH/rim-21	yes	no
<input type="checkbox"/>	FOIG_05151	<i>Fusarium oxysporum</i> f. sp. cubense tropical race 4 54006	unspecified product	yes	no
<input type="checkbox"/>	FOMQ_08199	<i>Fusarium oxysporum</i> f. sp. melonis 26406	hypothetical protein	yes	no
<input type="checkbox"/>	FOXG_09368	<i>Fusarium oxysporum</i> f. sp. lycopersici 4287	hypothetical protein	yes	no
<input type="checkbox"/>	FOZQ_12665	<i>Fusarium oxysporum</i> Fo47	hypothetical protein	yes	no
<input type="checkbox"/>	FVEG_06972	<i>Fusarium verticillioides</i> 7600	hypothetical protein	yes	no

Run clustalW for selected genes Check All Uncheck All
Select all orthologs or individual genes for ClustalW alignment

Orthologs and Paralogs within EuPathDB Data sets
Showing 9 of 69 rows
Click to run ClustalW
Fusarium

Protein clustalW	Gene	Organism	Product	Is syntenic	has comments
<input checked="" type="checkbox"/>	FFUJ_08659	<i>Fusarium fujikuroi</i> IMI 58289	related to pH signal transduction protein PalH	yes	no
<input checked="" type="checkbox"/>	FGRAAPH1_01G18125	<i>Fusarium graminearum</i> PH-1	pH-response regulator protein palH/rim-21	yes	no
<input checked="" type="checkbox"/>	FOCI_g10008702	<i>Fusarium oxysporum</i> f. sp. cubense race 1	pH-response regulator protein palH/rim-21	yes	no
<input checked="" type="checkbox"/>	FOCA_g10008637	<i>Fusarium oxysporum</i> f. sp. cubense race 4	pH-response regulator protein palH/rim-21	yes	no
<input checked="" type="checkbox"/>	FOIG_05151	<i>Fusarium oxysporum</i> f. sp. cubense tropical race 4 54006	unspecified product	yes	no
<input checked="" type="checkbox"/>	FOMQ_08199	<i>Fusarium oxysporum</i> f. sp. melonis 26406	hypothetical protein	yes	no
<input checked="" type="checkbox"/>	FOXG_09368	<i>Fusarium oxysporum</i> f. sp. lycopersici 4287	hypothetical protein	yes	no
<input checked="" type="checkbox"/>	FOZQ_12665	<i>Fusarium oxysporum</i> Fo47	hypothetical protein	yes	no
<input checked="" type="checkbox"/>	FVEG_06972	<i>Fusarium verticillioides</i> 7600	hypothetical protein	yes	no

Run clustalW for selected genes Check All Uncheck All

Note: You can also scroll down and select orthologs for multiple species without using the search box.

▼ Orthologs and Paralogs within EuPathDB [Data sets](#)

Search this table... Showing 69 rows

Protein clustalW	Gene	Organism	Product	is syntenic	has comments
<input type="checkbox"/>	Bcin08g03640	Botrytis cinerea B05.10	unspecified product	yes	no
<input checked="" type="checkbox"/>	FFUJ_08659	Fusarium fujikuroi IMI 58289	related to pH signal transduction protein PaliH	yes	no
<input checked="" type="checkbox"/>	FGRAMPH1_01T18125	Fusarium graminearum PH-1	ph-response regulator palH-rim-21	yes	no
<input checked="" type="checkbox"/>	FOC1_g10008702	Fusarium oxysporum f. sp. cubense race 1	pH-response regulator protein palH/rim-21	yes	no
<input checked="" type="checkbox"/>	FOC4_g10008637	Fusarium oxysporum f. sp. cubense race 4	pH-response regulator protein palH/rim-21	yes	no

3. Examine ClustalW output

CLUSTAL 2.1 Multiple Sequence Alignments

FOC1_g10008702-t38_1-p1	1MAAA	AALDMVSATI	TALPSLC	TTT	LLPDSG	VMTL	GLDGV	PVPLT	EPAVY	QVPCP	TPDLL	PRAEY	SVTVA	AASGDE
FOXG_09368-t26_1-p1	1MAAA	AALDMVSATI	TALPSLC	TTT	LLPDSG	VMTL	GLDGV	PVPLT	EPAVY	QVPCP	TPDLL	PRAEY	SVTVA	AASGDE
FOMG_08199-t38_1-p1	1MAAA	AALDMVSATI	TALPSLC	TTT	LLPDSG	VMTL	GLDGV	PVPLT	EPAVY	QVPCP	TPDLL	PRAEY	SVTVA	AASGDE
FOZG_12665-t36_1-p1	1MAAA	AALDMVSATI	TALPSLC	TTT	LLPDSG	VMTL	GLDGV	PVPLT	EPAVY	QVPCP	TPDLL	PRAEY	SVTVA	AASGDE
FOC4_g10008637-t38_1-p1	1MAAA	AALGVVSTTI	TALPSLC	TTT	LLPDSG	VMTL	GFDGV	PVPLT	EPAVY	QVPCP	TPDLL	PRAEY	SVTVA	AASGDE
FOIG_05151-t36_1-p1	1MAAA	AALGVVSTTI	TALPSLC	TTT	LLPDSG	VMTL	GFDGV	PVPLT	EPAVY	QVPCP	TPDLL	PRAEY	SVTVA	AASGDE
FFUJ_08659-t31_1-p1	1MAAA	AALDMVSATI	TGLPSLC	TTT	LLPDSG	VMTL	GFDGV	LVPLT	EPAVY	QVPCP	TPDLL	PRAEY	SVTVA	AASGDE
FVEG_06972-t26_1-p1	1MAAA	AALDMASATI	TGLPSLC	TTT	LLPDSG	IMTL	GFDGV	PVPLT	EPAVY	QVPCP	TPDLL	PRAEY	SVTVA	AASGDE
FGRAMPH1_01T18125-p1	1M	TAAGLISATI	TAPPSLC	TTT	LLPD	TGVMTL	GLDGV	PTILT	EPAVY	QIPCP	THDIL	PRAEV	HVTYA	AASGDE
NCU00007-t26_1-p1	1	MEPRQL	FSDP	TADPI	STAGA	ASNAL	SCASF	NLPEG	GILQL	-PNGE	IITLS	APAAF	KPPSC	NLALR	SVENI
FOC1_g10008702-t38_1-p1	75	SSLTRH	DPA	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
FOXG_09368-t26_1-p1	75	SSLTRH	DPA	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
FOMG_08199-t38_1-p1	75	SSLTRH	DPA	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
FOZG_12665-t36_1-p1	75	SSLTRH	DPA	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
FOC4_g10008637-t38_1-p1	75	SSLTRH	DPA	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
FOIG_05151-t36_1-p1	75	SSLTRH	DPA	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
FFUJ_08659-t31_1-p1	75	SSLTQH	DAD	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
FVEG_06972-t26_1-p1	75	SSLTQH	DASD	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
FGRAMPH1_01T18125-p1	72	SSLTQH	ELSD	SSPKF	SDFRD	PFYAS	TFPIC	YALAAT	TVTA	YMLVI	MLFVT	PRSF	LDGGIV	YLGRR	SAFTH
NCU00007-t26_1-p1	80	ASGT	MGLKAD	DDSH	FSDWRD	PFYAS	TFPQC	YALAAT	TVTA	YTLVI	MLFVT	PRSF	LDGGIV	VLGRK	-GFTN
FOC1_g10008702-t38_1-p1	155	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	HAAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
FOXG_09368-t26_1-p1	155	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	HAAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
FOMG_08199-t38_1-p1	155	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	HAAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
FOZG_12665-t36_1-p1	155	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	HAAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
FOC4_g10008637-t38_1-p1	155	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	HAAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
FOIG_05151-t36_1-p1	155	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	HAAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
FFUJ_08659-t31_1-p1	155	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	HAAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
FVEG_06972-t26_1-p1	155	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	HAAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
FGRAMPH1_01T18125-p1	152	RPWLQ	KVAAL	TVAI	SLTIAS	ADTF	FRVAKSQ	YMWGI	QNA	NQ	LQDE	VMNSVE	LKVIR	LVSDT	FL-----
NCU00007-t26_1-p1	158	RPWLQ	KVAAL	SVAI	SLTIAN	AATF	RAAEQQ	YSWG	VQNA	KQ	LOED	VLGGAE	LKIIR	IISDT	FLWLAQAQTL

```
.dnd file
(
(
(
(
(
FFUJ_08659-t31_1-p1:0.01996,
(
(
FGRAMPH1_01T18125-p1:0.08781,
NCU00007-t26_1-p1:0.47646)
:0.05084,
FVEG_06972-t26_1-p1:0.02306)
:0.00057)
:0.01607,
(
FOC4_g10008637-t38_1-p1:0.00116,
FOIG_05151-t36_1-p1:0.00013)
:0.00767)
:0.00231,
FOZG_12665-t36_1-p1:0.00082)
:0.00037,
FOMG_08199-t38_1-p1:-0.00033)
:0.00033,
FOC1_g10008702-t38_1-p1:0.00437,
FOXG_09368-t26_1-p1:-0.00033);
```

4. Use .dnd output to build a phylogeny tree

There are several web applications that can help you building a phylogeny tree for your ClustalW output. Here we will use **iTOL** as an example.

- Navigate to iTOL: <https://itol.embl.de/upload.cgi> in a separate tab
- Copy the **text** portion of the .dnd file from your ClustalW output tab

```
(
(
(
(
(
FFUJ_08659-t31_1-p1:0.01996,
(
(
FGRAMPH1_01T18125-p1:0.08781,
NCU00007-t26_1-p1:0.47646)
:0.05084,
FVEG_06972-t26_1-p1:0.02306)
:0.00057)
:0.01607,
(
FOC4_g10008637-t38_1-p1:0.00116,
FOIG_05151-t36_1-p1:0.00013)
:0.00767)
:0.00231,
FOZG_12665-t36_1-p1:0.00082)
:0.00037,
FOMG_08199-t38_1-p1:-0.00033)
:0.00033,
FOC1_g10008702-t38_1-p1:0.00437,
FOXG_09368-t26_1-p1:-0.00033);
```

- Paste the text into the **Tree text** section and click on the **Upload** button

Upload a new tree

Tree name:
optional

Paste your tree into the box below, or select a file using the **Tree file** selector. You can also simply drag and drop the tree file onto the page (only a regular plain text file, not QIIME QZA files).

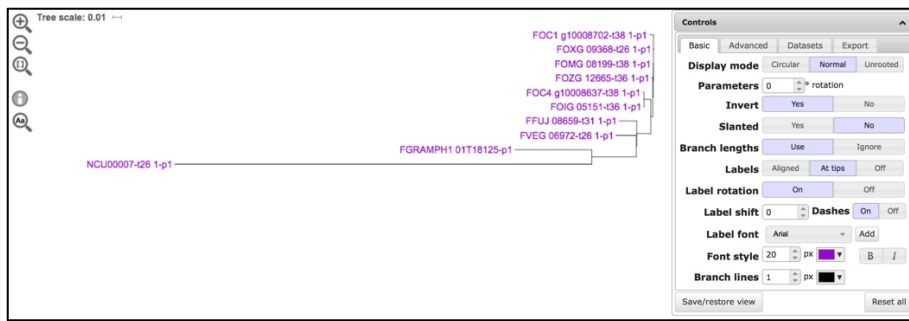
Tree text:

```
(
(
(
(
(
FFUJ_08659-t31_1-p1:0.01996,
(
(
FGRAMPH1_01T18125-p1:0.08781,
NCU00007-t26_1-p1:0.47646)
:0.05084,
```

Tree file:
Choose File no file selected

Upload

- Visualize an iTOL tree output and modify settings as needed



Useful links:

1. Useful information about ClustalW:
<https://www.ebi.ac.uk/Tools/msa/clustalw2/help/faq.html>
2. iTOL video tutorials:
https://itol.embl.de/video_tutorial.cgi