

BLAST databases

PTA clustering of public *Physcomitrella* sequences

ppp_raw	n P.p. public data, filtered raw sequences
ppp	n P.p. public data, PTA clustering
ppp_nr	n non-redundant representation of ppp
ppp_orfpep	p predicted ORFs from ppp_nr

provided on www.cosmoss.org

Retrieval **FAQ**
About **BLAST**
Links Documentation

Physcomitrella patens

www.
COSMOSS
.org



www.plant-biotech.net

Description of PTA -> FastA headers I/III

contigs

PTA clustering project PPP, cluster 375, contig 6 contig was build from 45 sequences
consensus length (contig) 1445 bp, contains 45 fragments, cluster contains 12 output
sequences of which 7 are contigs, cluster contains 86 input sequences, contig has ID #6

ppp_nr

```
>PPP_375_C6 consLen=1445 contFrag=45 contigs=7/12 clusterFrag=86 contigID=6 Members="pp001060  
093r, pp004055241r, gi|18367219 gb|BJ199300.1|BJ199300, gi|18374834|gb|BJ206412.1|BJ206412,  
gi|18368848|gb  
GGGGGCGCCGCTCTAGAACAGTGGATCCCCGGGCTGCAGGAATTCGGCACGAGCCATCTTCTCTGTTTTTTTTCCCCATT...
```

ppp

```
>PPP_375_C6 nseq=45  
GGGGGCGCCGCTCTAGAACAGTGGATCCCCGGGCTGCAGGAATTCGGCACGAGCCATCTTCTCTGTTTTTTTTCCCCATT...
```

ppp_orfpep

```
>PPP_375_C6 166.1 2096 398 117 1311 consLen=1445 contFrag=45 contigs=7/12 clusterFrag=86 con  
tigID=6 Members="pp001060093r, pp004055241r, gi|18367219 gb|BJ199300.1|BJ199300, gi|18374834  
|gb|BJ206412.1  
RLISNGDDHGRYCCRERFSSFAKLSSSAVAGAGVQNAAFPRSAAATSLGGKASSLYGASM...
```

ESTscan ORF prediction: norm_score raw_score cutoff start stop



Description of PTA -> FastA headers II/III

>project_cluster_contig contig was build from X sequences consensus length (contig), contains X fragments, cluster contains X output sequences of which X are contigs, cluster contains X input sequences, contig has ID #X clusters as members -> iterative

ppp_nr - contig

```
>PPP_375_C6 consLen=1445 contFrag=45 contigs=7/12 clusterFrag=86 contigID=6 Members="pp001060
093r, pp004055241r, gi|18367219 gb|BJ199300.1|BJ199300, gi|18374834|gb|BJ206412.1|BJ206412,
gi|18368848|gb
GGGGGCGCCGCTCTAGAACAGTGGATCCCCGGGCTGCAGGAATTCGGCACGAGCCATCTTCTCTGTTTTTTTTCCCCATT...
```

ppp_nr - contig from iterative cluster

```
>PPP_42_CB5 consLen=2700 contFrag=4 contigs=18/34 clusterFrag=193 contigID=5 Members="PPP.42.C
17, PPP.42.C9, PPP.42.C16, PPP.42.C23"
CCGCGTTGAACAGCTTACGTTTCGGCCGCGCCAGCAGCTTGCTCATGATCGTGGACGTGAACAGCGCCGCGTGCTTGATC...
```

ppp_nr - seed cluster contig

```
>PPP_sd_90_C1 consLen=765 contFrag=4 contigs=1/1 clusterFrag=4 contigID=1 Members="pp00600200
9r, gi|18345399 gb|BJ177442.1|BJ177442, pp004088150r, gi|22759857 dbj|AB048210.1| Physcomitr
ella patens Re
TCGCTGCAGGTCGAGACTACTGGATCCAAAGAATTCGGCACGAGGCTGCTATGTGTGTCGAAGCAGTGCTTGTAGTTGATG...
```

○ seed cluster -> >project_sd_seed-cluster_contig



Description of PTA -> FastA headers III/III

singlets from clusters

ppp_nr

```
>BJ157137 gb|BJ157137.1|BJ157137 consLen=630 contFrag=1 contigs=0/2 clusterFrag=2  
contigID=- cluster=PPP_7080  
CTTGTGTCAGCTTTGTGGGTGAGCTCGCAGCTGTGATTACTCTTGCCATCATGGAAACCAGTGGGTATGTGTGTCTAACT...
```

ppp

```
>BJ147381 PPP_cl.425.singlet  
CCAGACGCGGCTCGTGGGCGGCGTCTAAACTATGGATCCCCGGGCTGAGGAATTCGGCAC...
```

clone id / genbank acc. nr.
cluster id

non-cluster singlet

ppp/ppp_nr

```
>BU149122  
cttaaccocctttattgaaagccctgcacgagttttgacatactttcttttcaactcacact...
```

problem sequence

ppp/ppp_nr

```
>PR-AW128834  
ccctcaagattactctgaagcacgagcccaactctgcgtttttgggatgcaagagtaagg...
```

