**List of Functions Used in Ortho\_Detection V-1\_Old**

1. **GetMatrixNumber()**

Receive the info about matrix Number and Return the Matrix name Like "\n1. BLOSUM45\n2. BLOSUM62\n3. BLOSUM82\n4. Quit". BLOcks Substitution Matrix (BLOSUM) is a Substitution matrix used for sequence alignment of Proteins. BLOSUm45 is for more distantly related Proteins alignment Database 2. BLOSUM62 : MidRange Seq with more than 62%s Similarity. 3. BLOSUM82 :- More Related Proteins

1. **GetQuerySequence(genome):**

Receive genome Fastaq File and return gene\_sequence\_list

1. **WriteQuery(query, Parallel\_num)**
2. **RunBlast(subject, Parallel\_num)**
3. **Get\_Same\_Species\_Forward\_Best\_Hit(blastp\_score)**
4. **GetForwardBesthit(blastp\_score)**
5. **DivisionParallelQuery(queryV,query\_division\_value, cpu\_count, queryV\_len)**
6. **RunParallelQuery(species\_of\_query , species\_of\_subject , queryV, parallel\_num)**
7. **Oneway\_Threshold\_Best\_hit(mode)**
8. **Backward\_Best\_Hit(args)**
9. **Search\_Equal\_BBH\_Data(target\_A)**
10. **Search\_Unequal\_BBH\_Data(target\_B)**
11. **Matching\_BBH(target)**
12. **Generating\_Matrix\_Clustering\_Ortholog(element\_set, bar)**
13. **Parallel\_MCL(score\_matrix)**
14. **MCL(score\_matrix)**
15. **Clustering(row\_data, col\_data, score\_matrix)**
16. **Parallel\_Matrix\_Multiplication\_Using\_Numpy(data)**
17. **Parallel\_Matrix\_Power\_Using\_Numpy(matrix\_element)**
18. **Parallel\_Matrix\_Divide\_Using\_Numpy(data)**
19. **Read\_Species\_List(pr=0)** This Function is Only to display the name of Species inside the species Folder and return selected\_species\_dic (dict) , backward\_selected\_species\_dic, and total files inside species folder

(selected\_species\_dic, backward\_selected\_species\_dic, number\_i)

1. **Del\_File(path, file)**
2. **Check\_File(File)**
3. **Read\_Equal\_BBH(path)**
4. **Read\_Unequal\_BBH(path)**
5. **Read\_species\_List()**

**Blastp\_score is output file produced by runblast()**

**Blastp\_score = RunBlast(selected\_species\_dic[Species\_of\_object, parallel\_num]**

**~~Parallel\_num is number of CPU~~**

**#Steps of Program to run**

1. **Read\_Species\_List(pr =1) Run first and return 3 values**
2. **GetMatrixNumber() 🡪 blastp\_matrix**
3. **Check\_file(Cluster\_out) # Cluster\_out = command\_options.Cluster\_out**
4. **Del\_File(Score\_file, “\*”) # Score\_File = command\_options.Score\_file**
5. **Starting Log file Write**
6. **Backward\_best\_hit\_work\_list = Oneway\_Threshold\_Best\_Hit(mode) # mode = command\_options.mode**
7. **Read\_Equal\_BBH(Score\_file+selected\_species\_dic[i]+”\_”+selected\_specied\_dic[k])**
8. **Read\_unequal\_BBH(Score\_file+selected\_species\_dic[i]+”\_”+selected\_species\_dic[k])**
9. **Matchin\_BBH(unequal\_RBH\_element)**
10. **Generating\_Matrix\_Clustering\_Ortholog(data, bar)**

**## variable Name with Running Process**

1. **Mode:. Is user input to select mode among 3 that is Blastp , Blastp Using precalculated data and Clustering**
2. **Second Variable is to select Name of Genome to analyze (Suppose Only one Selected)**

**Genome\_name selected is passed to selected\_number and converted to sorted set and also program exit if input greater than number of Genome available**

1. **Blastp\_matrix = GetMatrix() Function run and Output is “BLOSUM45” , “BLOSUM62” or “BLOSUM” default is BLOSUM62**

**BLOSUM80 is used for closely related alignment databases, BLOSUM45 is used for more distantly related alignments BLOSUM62 is the Matrix built using sequences with more than 62 % Similarity (Sequences with >= 62 %identically were clustered. Note BLOSUM62 is the default matrix for protein BLAST. Experimentation has shown that the BLOSUM-62 matrix is among the best for detecting most weak protein similarities.**

**user\_selected\_number = [backward\_selected\_species\_dic[ele] for ele in genomes]**

**If mode 1 Passed**

1. **Backward\_best\_hit\_work\_list = oneway\_Threshold\_Best\_Hit(mode)**
2. **Pool = multiprocessing.Pool(cpu\_count)**

**Description of All Variables**

1. **Mode** = User input to select Mode (1 is for Blastp. 2. Blastp using precalculated data 3. Clustering

**1**

**Selected\_species\_dic** = Dictionary value of all species inside species Folder **1**

1. **Backward\_selected\_species\_dic** = Same as 2 But the Dictionary is Opposite

**Number\_i** = length of File inside species Folder **4**

1. **Selected\_number** = Number to select Gene like 1.AAE Gene Position in the Folder. This value is Used to take Out Species
2. **cpu\_count** = The Value of CPU For Parallel Counting (For Our Training We Only Use 1) \
3. **score\_file** = command\_options.score\_file (User input File)