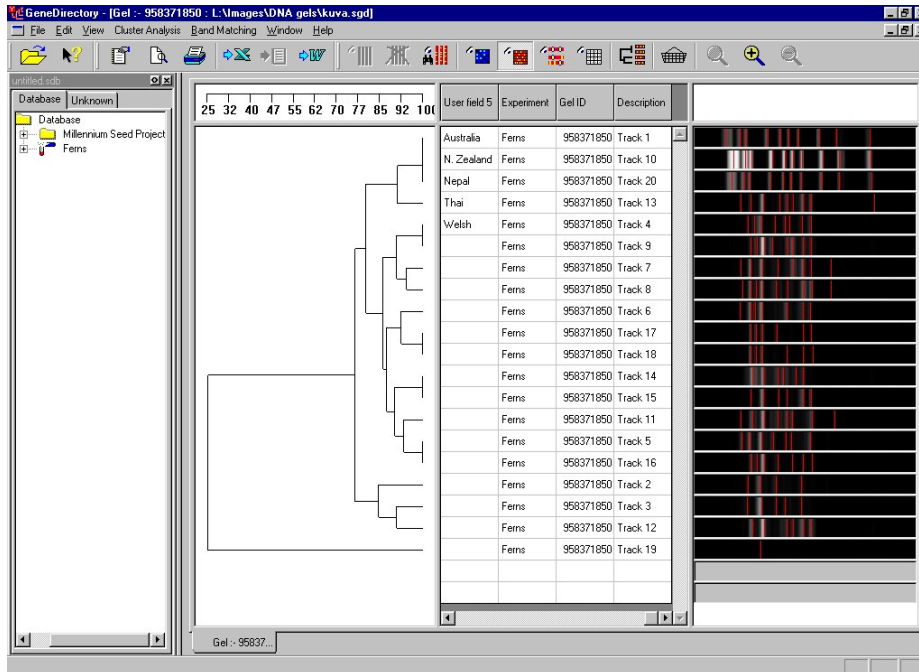# GeneDirectory Demonstration Notes

Guide 1

SYNGENE

Syngene
A Division of the Synoptics Group
Beacon House
Nuffield Road
Cambridge
CB4 1TF

Tel: +44 (0)1223 727123
Fax: +44 (0)1223 727101
www.syngene.com

GeneDirectory Demonstration Guide 1

Cluster Analysis – Similarity Matrix and Dendrogram, Single Reference Track and Match to all Tracks

Scope
Creating and saving a database in GeneDirectory
Connecting and Archive an Experiment (gel(s) containing tracks) from GeneTools to GeneDirectory
Editing and entering Classification data for the Tracks in the Experiment
Generating a Band Matching analysis/Cluster Analysis in GeneTools and GeneDirectory Dendrogram
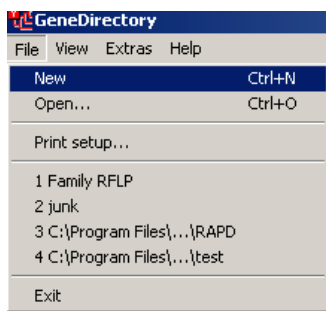Calculating Dendrogram distances
Similarity Matrix
Comparing a Cluster Analysis based on a single reference track and matching to all tracks

Demo Image - DemoMatch

Creating and saving a database in GeneDirectory
1. Open GeneDirectory.
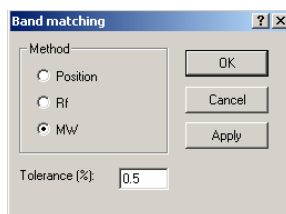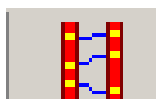2. Go to File/New to create and save a new database.

3. Open GeneTools and open the image DemoMatch.
4. Note Tracks 7 and 9 are denoted molecular weight standard tracks, that is, have molecular weight assigned to the bands in each Track.
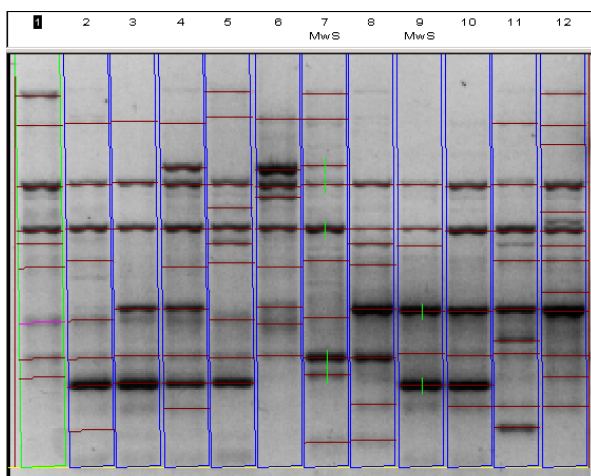
5. Highlight Track 7 as the Matching Standard (right click and select on the Track).
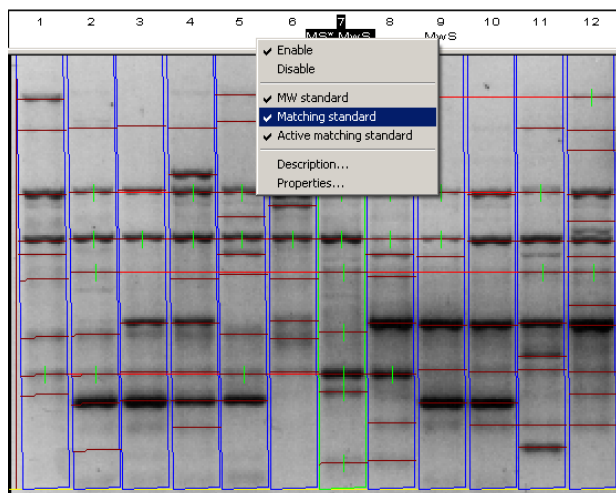
6. Select the Band Matching icon

and choose the fields Mw for the Method and 0.5 for the % Tolerance. Click 'Apply' and 'Ok'.

Connecting and Archive an Experiment from GeneTools to GeneDirectory
7. Click the Connect to Database icon and browse for the Database created and saved in step 2.


8. Click the Archive to Database icon in GeneTools.

9. Check 'All Tracks including the molecular weight standard tracks'; in the Archive to Experiment tab ensure the 'Database' folder is highlighted and select 'New Experiment'.
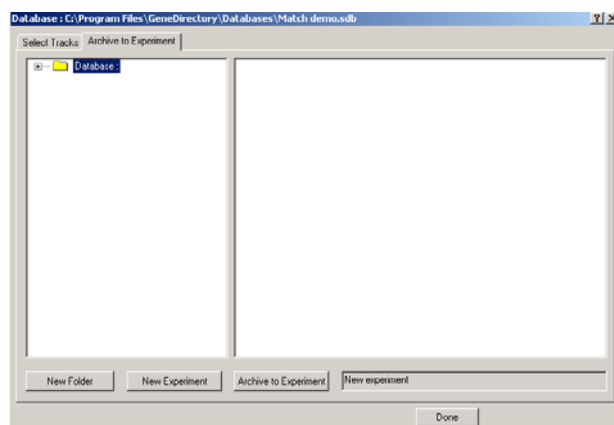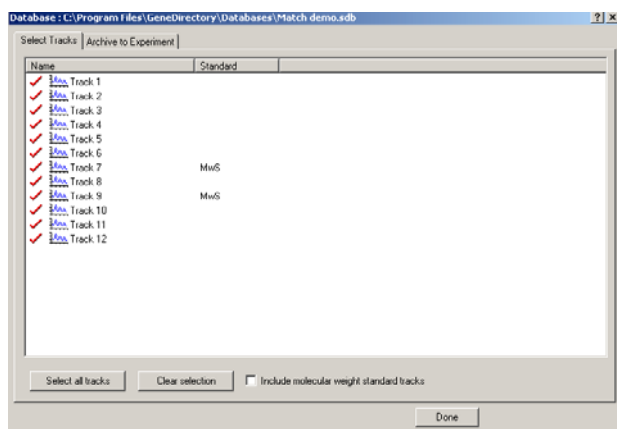
10. A New Experiment folder should appear under the Database folder. If it is not visible expand the tree by clicking on +. Rename to 'Experiment Match gel 1'
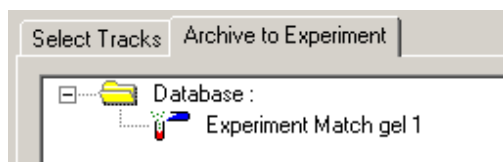
11. Click 'Archive to Experiment'. The experiment should now appear displayed in the right display. Select 'Done'. The experiment (gel containing tracks) has now been imported into the GeneDirectory database.
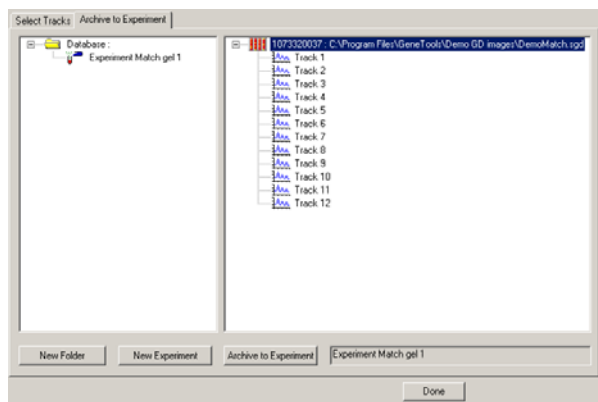
12. Maximise GeneDirectory (you will be in the explorer view by default) the Database tab will contain your archived experiment. It will also appear in the Unknown tab as no classification has been undertaken. If not visible then hit F5 (refresh).

Editing and entering Classification data for the Tracks in the Experiment

13. Go to Edit/Data Headings and delete 2 of the Headings (Columns 4 and 5) and re-name (Alter) 3 of the headings - Heading 1 to 'Genus', Heading 2 to 'Species' and Heading 3 to 'Strain'.
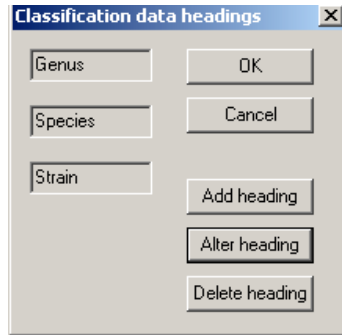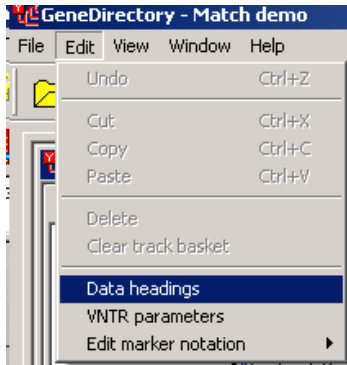
Generating a Band Matching analysis/Cluster Analysis in GeneDirectory

14. Double click on 'Experiment Match gel 1' experiment in the explorer view, which will result in a new display. There will be 3 vertical panes: blank, Track List and Tracks. Resize these panes appropriately.
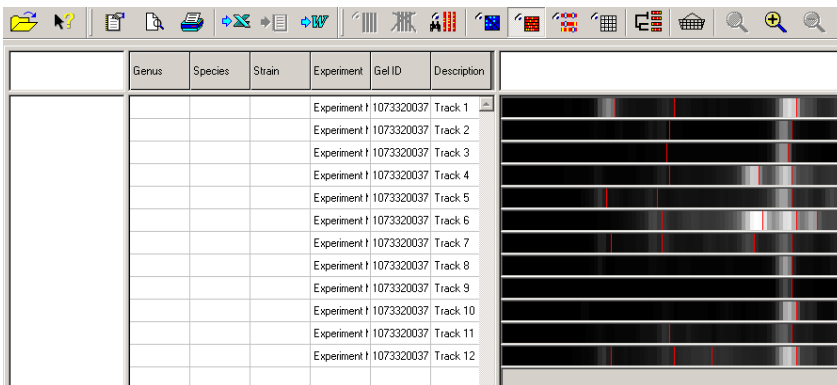
15. Double click on the first Track in the Track List pane and type in 'E for' Genus, 'Coli' for Species and 'K12' for Strain. Click 'Ok'.
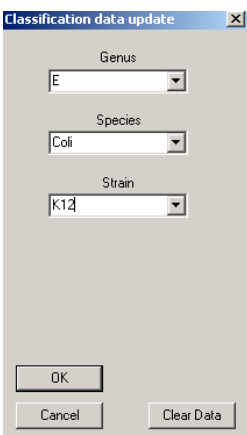
16. Double click on the next Track in the Track list pane and from the 3 drop down menus select "E, Coli and K12". Repeat for the remaining Tracks.

17. Right click on Track 7 in the Track List pane and select Reference Track.

18. Go to Cluster Analysis/Calculate selecting the following: Match to Single Reference, Mw, 0.5%, Dice, Neighbor Joining. Click 'Apply' and 'Ok'.

Dendrogram

19. A Dendrogram should appear in the first vertical pane of the display (identical to the Dendrogram in GeneTools). To view the Dendrogram distance values go to View/Dendrogram Labels.

20. Return to GeneTools and in the results table click on the Matching Matrix tab where a cross denotes a match to the Reference Track. Go to File/Export table to Excel. The results table will now have been imported into Excel and each cross will have been replaced by a '1'. Within this Excel spreadsheet enter a total for each Track. For instance the totals for Tracks 7, 2 and 5 are 10, 4 and 3, respectively.

| Results for selected track | Results for all tracks | Matching comparisons | **Matching matrix** | Matching coeffic |

Peak	Track 5	Track 6	Track 7	Track 8	Track 9	Track 10	Track 11	
1			✕					
2		✕	✕					
3			✕					
4	✕		✕	✕	✕		✕	
5	✕	✕	✕	✕	✕			
6			✕				✕	
7			✕					
8	✕		✕	✕				
9			✕					
10			✕					

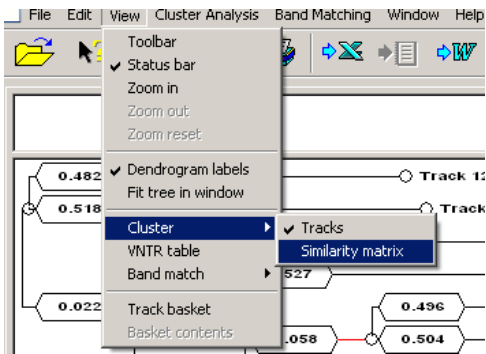|	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Peak	Track 1	Track 2	Track 3	Track 4	Track 5	Track 6	Track 7	Track 8	Track 9	Track 10	Track 11	Track 1;
2	1							1					
3	2						1	1					
4	3							1					
5	4		1		1	1		1	1	1		1	
6	5		1	1	1	1	1	1	1	1			
7	6		1					1				1	
8	7							1					
9	8	1	1			1		1	1				
10	9							1					
11	10							1					
12		1	4	1	2	3	2	10	3	2	0	2	

Calculating Dendrogram distances

21. Returning to GeneDirectory observe on the Dendrogram the following clusters of 2 Tracks only: 7 and 2, 5 and 8, 9 and 4.
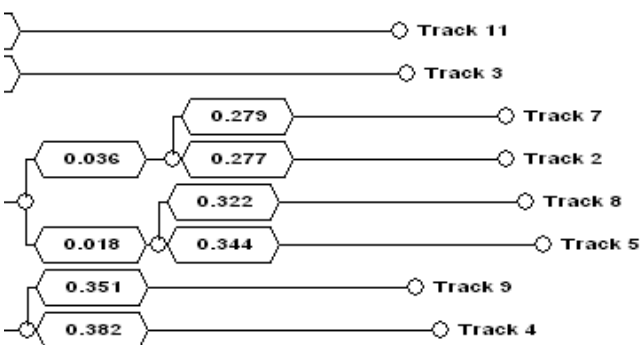
22. The distance between 2 Tracks indicates the degree of similarity so the shorter the distance, the more similar. The approximate distance can be calculated by summing the values between the Tracks. For example Track 7 to Track 2 has a distance of 0.279+ 0.277 =0.556.

Similarity Matrix

23. The degree of similarity can also be seen in the Similarity matrix (go to View/Cluster/Similarity Matrix).

24. The closer the number to 1 the more similar the Tracks. In this case for Track 7 and 2 the similarity coefficient is 0.44 (the sum of the dendrogram distance and the similarity coefficient is approximately 1). If you now review the Excel spreadsheet in GeneTools Tracks 7 and 2 have 4 bands in common. The Dendrogram is a graphical representation of the similarity of the Tracks and is a preferred view to the Similarity Matrix and Matching Matrix as the relationship within and between clusters can also be seen.

25. Tracks 5 and 8 are clustered, the distance being 0.322+0.344= 0.666. The number of bands that match with respect to Track 7 is 3. Tracks 9 and 4 are clustered and distance value is 0.351+0.382=0.733. The number of bands in common with respect to Track 7 is 2. Summary of the cluster pair results:

Cluster pair	No. bands matching (Reference Track 7)	Dendrogram distance	Similarity Coefficient
7 and 2	4	0.556	0.440
5 and 8	3	0.666	0.330
9 and 4	2	0.733	0.260

Comparing a Cluster Analysis based on a single reference track and matching to all tracks

26. The Matching/Similarity coefficients and Dendrogram views can be seen in both GeneTools and GeneDirectory, the only difference being the display views. For the example chosen, we have matched to a Single Reference Track (Track 7). In some cases matching all Tracks to all Tracks is needed and this can only be performed in GeneDirectory.

27. In GeneDirectory go to Cluster Analysis/Calculate and leaving all the fields the same, change to Matching All Tracks to All Tracks. Click 'Apply ' and 'Ok'.
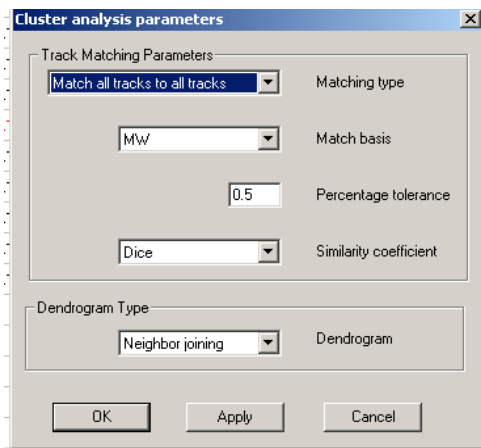
28. Both the Similarity Matrix and Dendrogram views will have changed. Generally the values in the Similarity matrix will be higher. When matching all tracks to all tracks, if one Track has 1 band and the Track it is being compared to has 2 bands, and there is 1 match, then both these Tracks are very similar to each other (1 match and 1 non match). However if you now compare these 2 Tracks to a Reference Track that contains 10 bands then the Tracks are now not very similar to the Reference Track (only 1 match and 9 non matches). This is clearly seen in the respective Dendrogram views. For example the distance was calculated for Track 3 to 4 for both matching to a single reference and to all matches:

Track 3 to 4	Match to Single Reference	Match all Tracks to all Tracks
Dendrogram distance	0.895	0.467
Similarity Coefficient	0.130	0.530
No. bands matching	1	1
No. bands not matching	9	1

NB Dendrogram distance calculations:
Single Reference: 0.382+0.029+0.038-0.058+0.504=0.895
Match to all Tracks: 0.184+0.283=0.467