

BIONUMERICS Tutorial:

Identifying unknown samples based on 16s rDNA sequences

1 Aim

BIONUMERICS contains powerful tools for the identification of unknown samples against a reference set. With the internal validation options, the user knows exactly how reliable the identification is and which type of errors can be expected. Different datatypes or combinations of datatypes can be used for identification. In this tutorial we will use the 16s sequences available in the **Demobase Connected**.

2 Preparing the database

The **DemoBase Connected** will be used in this tutorial and can be downloaded directly from the *BIONUMERICS Startup* window or restored from the back-up file available on our website:

- 1. To download the database directly from the *BIONUMERICS Startup* window, click the button, located in the toolbar in the *BIONUMERICS Startup* window. Select **DemoBase Connected** from the list and select **Database** > **Download** (). Confirm the download action.
- 2. To restore the database from the back-up file, first download the file DemoBase_Connected.bnbk from https://www.applied-maths.com/download/sample-data, under 'DemoBase Con-

nected'. In the *BIONUMERICS Startup* window, press the b button, select **Restore** *database*, browse for the downloaded file and select **Create copy**. Specify a name and click < OK >.



In contrast to other browsers, some versions of Internet Explorer rename the DemoBase_Connected.bnbk database backup file into DemoBase_Connected.zip. If this happens, you should manually remove the .zip file extension and replace with .bnbk. A warning will appear ("If you change a file name extension, the file might become unusable."), but you can safely confirm this action. Keep in mind that Windows might not display the .zip file extension if the option "Hide extensions for known file types" is checked in your Windows folder options.

3 Creating the reference comparison

Before creating an identification project, we will first create a comparison containing the reference set against which our unknown samples will be identified.

1. In the *BIONUMERICS Startup* window, double-click on the **DemoBase Connected** database to open it.

Select all samples with an identification down to species level. This can be done as follows:

- 2. Selecting *Edit* > *Search...* (Ctrl+F) to open the *Find objects in view* dialog box (see Figure 1).
- 3. Select the *Genus* row in the grid, and specify "STANDARD" in the text box. Press the *NOT* button in the *Group by* panel.
- 4. Select the *Species* row in the grid, and specify "sp." in the text box. Press the *<NOT>* button in the *Group by* panel.

Two rules are now defined in the query, one that the field **Genus** should not contain **STANDARD** and a second that the field **Species** should not contain **sp.** (see Figure 1).

Find objects in view	? ×
Entries	Statements
Key	Add new
Modified date	Remove selected
NOT Genus Contains	Remove all unused
NOT Species Contains sp.	Include all visible fields
Strain number	Group by AND OR NOT Reset query Add to selection
< >	Show selection view

Figure 1: The *Find objects in view* dialog box with the query to select all samples identified down to species level.

5. Press <*OK*>.

If the query is run correctly, 42 entries are now selected. This is our reference set.

Highlight the *Comparisons* panel in the *Main* window and select *Edit* > *Create new object...* (+) to create a new comparison with our reference set.

A cluster analysis can be performed using the 16s sequences to evaluate how well separated the data is. This is not necessary to perform the identification though it can affect some of the options

available in the identification project.

Select 16S rDNA in the *Experiments* panel, select *Clustering* > *Calculate* > *Cluster anal-ysis (similarity matrix)...*, use the similarity coefficient based on the *Standard* pairwise alignment, press < *Next* > and <*Finish* >.

Now we will create groups that reflect the Genus and Species of each sample:

- Right-click on the column header of the field *Genus* in the *Information fields* panel and select *Groups* > *Create groups from database field*, leave the settings at default and press < OK >.
- 9. Repeat previous action for the field **Species**, but this time select the option to subdivide the existing groups (see Figure 2).

Create groups from f ? ×
Group creation priority
Create largest groups first
 Create groups alphabetically
Create groups by occurrence
Subdivide existing groups
Skip empty content
Maximum count 30
OK Cancel

Figure 2: Subdivide existing groups.

There are now six groups present in the comparison (see Figure 3).

10. Save the comparison with *File* > *Save* (B, Ctrl+S), name it **RefSet** and close it with *File* > *Exit*.

dit Layout Groups Cluste 局 ि ↓≭¼ £ी				ReadSets Spe		site Window Help						
riments								_				
Experiment types>	0		Dendrogram			Experiment data	Information fields				Similarities	
Name	Aspect		이 바 영 양	데 / 1			8.11.13↑↓₩	† ↓			i ↓†∔ III	
RFLP1	<all band="" classes=""></all>	•		L* L		16S rDNA						
RFLP2	<all band="" classes=""></all>											
AFLP											16S rDNA 0 20 40 60 80	100
	<all band="" classes=""></all>		16S rDNA 82.147%	66	100		Group	Key	Genus		0 20 40 60 80	100
PhenoTest	<all characters=""></all>		85 90	· · · · "		TCTCTAAATCAGGA		G@Gel08@014	Ambiorix	sylv		
FAME	<all characters=""></all>				p[TCTCTAAATCAGGA	Ambiorix,sylvestris	G@Gel09@011	Ambiorix	sylv		
16S rDNA	<defaut></defaut>				- F	TCTCTAAATCAGGA'	Ambiorix, sylvestris	G@Gel07@005	Ambiorix	sylv		
DNA-Hybrid	<defaut></defaut>				Ę	CCCGGATTGAGCTT	Ambiorix, sylvestris Ambiorix, sylvestris Ambiorix, sylvestris Ambiorix, sylvestris Ambiorix, sylvestris	G@Gel07@007 G@Gel09@010	Ambiorix	sylv sylv		
- /		~			-	TCTCTAAATCAGGA'	Ambiorix, sylvestris	G@Gel07@011	Ambiorix	sylv		
					E.	TCTCTAAATCAGGA'	Ambiorix,sylvestris Ambiorix,sylvestris	G@Gel07@013 G@Gel07@014	Ambiorix	sylv sylv		
es Local composite datasets						TCTCTAAATCAGGA	Ambiorix, sylvestris	G@Gel11@008	Ambiorix	sylv		
\otimes						TCTCTAAATCAGGA	Ambiorix,sylvestris Ambiorix,sylvestris Ambiorix,sylvestris Ambiorix,sylvestris	G@Gel07@012	Ambiorix	sylv		
ame					ſĽ	TCTCTAAATCAGGA'	Ambiorix,sylvestris Ambiorix,sylvestris	G@Gel08@010 G@Gel08@013	Ambiorix Ambiorix	sylv sylv		
6S rDNA						CTGTTGTTCGTTTA	Ambiorix, sylvestris Ambiorix, sylvestris Ambiorix, sylvestris Ambiorix, sylvestris Ambiorix, sylvestris	G@Gel07@002	Ambiortx	sylv		
65 IUNA		_			П -г	TTGTATGAACCGCG	Ambiorix, sylvestris	G@Gel11@003	Ambiorix	sylv		
						TCTAAATCAGGATT: ATCAGGATTGAGCT	Ambiorix,sylvestris Ambiorix,sylvestris	G@Gel11@009 G@Gel09@004	Ambiorix Ambiorix	sylv sylv		
					r	TCTCTAAATCAGGA'	Ambiorix, aberrans	G@Gel07@003	Ambiortx	abe		•
					- L	TCTCTAAATCAGGA' TCTCTAAATCAGGA'	Ambiorix, aberrans	G@Gel11@010	Ambiorix	abe		
		- 18				TCTCTAAATCAGGA	Ambiorix, aberrans Ambiorix, aberrans	G@Gel07@016 G@Gel11@004	Ambiorix Ambiorix	abe		
		- 1			r	TCTCTAAATCAGGA	Perdrix,pseudoarchaeus	G@Gel08@005	Perdrix	pse		
08					L.	CAGGATTGAGCTTC	Perdrix,pseudoarchaeus Perdrix,pseudoarchaeus	G@Gel09@009	Perdrix	pse		
					1	TCTCTAAATCAGGA'	Perdrix,pseudoarchaeus Perdrix,pseudoarchaeus	G@Gel09@002 G@Gel09@005	Perdrix	pse		
⊗ 誓 耆 ≌>				-	L	TCTCTAAATCAGGA	Perdrix, pseudoarchaeus	G@Gel09@007	Perdrix	pse		
		-			-	CAGGATTGAGCTTC: ACGGATTGAGCTTC:	Perdrix,pseudoarchaeus	G@Gel08@002	Perdrix	pse		
16 Ambiorix, sylvestris					Æ	TGATTGTTGTATGA.	Perdrix,pseudoarchaeus Perdrix,pseudoarchaeus	G@Gel08@004 G@Gel08@007	Perdrix	pse		
14 Perdrix,pseudoarchaeus					L.	TCTCTAAATCAGGA'	Perdrix.pseudoarchaeus	G@Gel09@008	Perdrix	pse		
4 Ambiorix, aberrans		- 1			-	TCTCTAAATCAGGA'	Perdrix, pseudoarchaeus	G@Gel08@015	Perdrix	pse		
3 Vercingetorix, palustris						TCTCTAAATCAGGA	Perdrix,pseudoarchaeus Perdrix,pseudoarchaeus	G@Gel09@003 G@Gel08@003	Perdrix Perdrix	pse pse		
3 Vercingetorix, nemorosum						ATTCTTGTATGAAC	Perdrix,pseudoarchaeus	G@Gel08@006	Perdrix	pse		
2 Vercingetorix, aquaticus						CCGGTGATTGTTGT.	Perdrix.pseudoarchaeus	G@Gel08@011	Perdrix	pse		
					Ъ	ATAAATTCTCTAAA'	Vercingetorix, nemorosum Vercingetorix, nemorosum Vercingetorix, nemorosum	G@Gel07@010	Vercingetorix	nem		
		- 11			L'	TTGTTGTATGAACC	Vercingetorix, nemorosum	G@Gel11@011 G@Gel11@006	Vercingetorix Vercingetorix	nem		
						TATGAACCGCGCAT	Vercingetorix, aquaticus	G@Gel07@015	Vercingetorix	aqu		
						TATGAACCGCGCAT	Vercingetorix, aquaticus	G@Gel07@017	Vercingetorix	aqu		
		- 11			Ъ	CGTCCGTGATTGTT(Vercingetorix, palustris Vercingetorix, palustris	G@Gel07@004 G@Gel08@016	Vercingetorix Vercingetorix	palu palu		
						GATTGTTGTATGAA	Vercingetorix, palustris Vercingetorix, palustris	G@Gel11@005	Vercingetorix	palu		

Figure 3: The Comparison window with groups defined.

4 Creating the identification project

The reference set is now ready to base our identification project on.

- 1. Highlight the *Identification projects* panel in the *Main* window and select *Edit* > *Create new object...* (+).
- 2. In the first step of the New identification project wizard, select the comparison RefSet and leave the option to lock the reference comparison checked (see Figure 4). This will safeguard the comparison against any accidental changes that might affect the identification results. Press < Next>.

lew identification project			?	×
Reference comparison Please select the comparison that	contains the re	ference data.		
<all comparisons=""></all>	~			
RefSet				
∠ Lock reference comparison				
Contraction Companison				
	< Bac	Next >	Са	ncel

Figure 4: Select reference data.

- If the reference comparison is locked, changes can only be applied if the comparison is unlocked first. To do this, open the comparison and select *File* > *Object access status...* to open the *Object access* dialog box. Check the radio button next to *Unlocked* and click < *OK* >. It is strongly recommended that after applying the desired changes the comparison is locked again.
 - 3. In the second window of *New identification project* wizard, make sure *Comparison groups* are checked as class labels and click <*Finish*> (see Figure 5).
 - 4. Optionally, you can change the name of the project. Press < OK >.

We have now defined where our reference set is and what we wish to use as label for the identification. Next, we need to define the experiment and the algorithm, this is stored in the classifier. Per identification project, several classifiers can be defined in order to compare identification results from different experiments and /or algorithms, but in this tutorial, we will only create one classifier.

5. Create a new classifier by selecting *Edit* > *Create new classifier...* (+) in the *Identification project* window.

This opens the New classifier wizard.

6. In the first step, select 16S rDNA and press < Next>.

ass labels Please select the source of the reference class labels. Comparison groups The comparison group name is used as class identifier. Information field(s) One or more entry information fields is used as class identifier. Information fields> Genus Species Strain number	×
The comparison group name is used as class identifier. Information field(s) One or more entry information fields is used as class identifier. It Entry fields> Genus Species	
Information field(s) One or more entry information fields is used as class identifier. Ill Entry fields>	
One or more entry information fields is used as class identifier.	
All Entry fields>	
Genus Species	
Species	
Strain number	
< Back Finish Car	ncel

Figure 5: Select reference class labels.

In the second step, all algorithms compatible with this experiment are listed. This means that this list is different for different experiment types.

- 7. Select the first method Similarity values (Standard) and click < Next>.
- 8. In the third step of the *New classifier* wizard, choose **Balanced similarity** as scoring method and press <*Next*>.
- 9. In the fourth step, define a threshold of 98 % similarity with a minimum difference of 1 and click <*Next*>.

Rank by score	Similarity		\sim		
-Reported sco	res				
	Threshold:	Min. difference:			
Similarity	98	1			

Figure 6: Specify thresholds.

10. You will be prompted for a name of the classifier, just leave the name at the default suggestion and click < OK >.

5 Validating a classifier

The classifier is now present in our identification project and ready for use. However, it is advised to optimise the classifier parameters and to run a validation on the classifier to check its performance before using it for identification purposes.

- 1. For determining the optimal parameters, select *Edit* > *Optimise classifier parameter...* (1/2) to open the *Classifier parameter optimization* dialog box.
- 2. Select the Max. similarity weight and click < OK > to start optimizing this parameter.

The calculation of the optimization may take several minutes, depending on the capacity of your PC. The result of the optimization is a maximum similarity weight of 0.85. This value will be automatically filled in, in the classifier.

Optimization	×
The optimal value for BF in the interval [0,1] is 0.85 This value is set in the classifier	
ОК	

Figure 7: Optimized parameter.

3. A tool for internal validation has been included in the software and can be run by selecting *Edit* > *Cross-validation analysis...* (☆). This will open the *Cross-validation* dialog box, leave the settings at default and click <*OK*>.



The validation analysis can take quite some time, especially on large reference sets. In these cases it is advised to increase the test group size and decrease the coverage.

After the cross validation has finished, a detailed overview of the results are shown (see Figure 8). Clicking on a cell in the confusion matrix will give a detailed overview on the entries in this cell in the lower right panel. Some predefined graphs are also available to get a better understanding on the performance of the classifier.

4. To open a predefined graph, select *Edit* > *Chart and statistics...* (h).

5. Choose the first graph **Cross validation class score contrasts** and click *<OK>*.

This graph (see Figure 9) compares the average reference score with the average best nonreference score, showing also the standard deviations. A good classifier shows a clear separation of these two scores for all groups. In this case, the contrast between these scores is highest for Perdrix pseudoarchaeus. This is to be expected as this is the only species belonging to genus Perdrix.



The reliability of the cross validation depends greatly on the reference set. For smaller sets or sets that do not contain all representative samples, the results of the cross validation should be interpreted with care. For example, in this set we only have one species of the genus Perdrix, the cross validation shows us how well it can be identified compared to the other five species in the dataset, but it does not provide information on how well Perdrix pseudoarchaeus is separated from other Perdrix species.

Identification cross valid								- 🗆
e Edit Window Help								
Count (relative)	↓	Δh.						
Confusion matrix								
Class name	Class total	True positives	False negatives	False positives	Unmatched	Ambiorix,sylves	Perdrix,pseudo	Ambiorix,aberrans
All classes>	42	95.24%	0.00%		4.76%			
Ambiorix, sylvestris	16	93.75%	0.00%	0.00%	6.25%	93.75%	0.00%	0.00%
erdrix,pseudoarchaeus	14	100.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%
Ambiorix, aberrans	4	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
/ercingetorix,palustris	3	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
/ercingetorix,nemorosum	3	66.67%	0.00%	0.00%	33.33%	0.00%	0.00%	0.00%
/ercingetorix,aquaticus	2	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	<							>
Settings			Entries					
			Ref. cla	nss Ref. o	ass score Ma	tch class 🛛 🛛	latch class score	Кеу
- Statistics			\sim					
Recall: 95.24%								
Precision: 100.00	104							
F1 measure: 97 5								
	50 /0							
 Classifier 								
Deference comp	arison: RefSet							
Reference comp								
Reference labels	: Comparison groups		\sim					

Figure 8: The *Identification cross validation* window with the results of the internal validation.

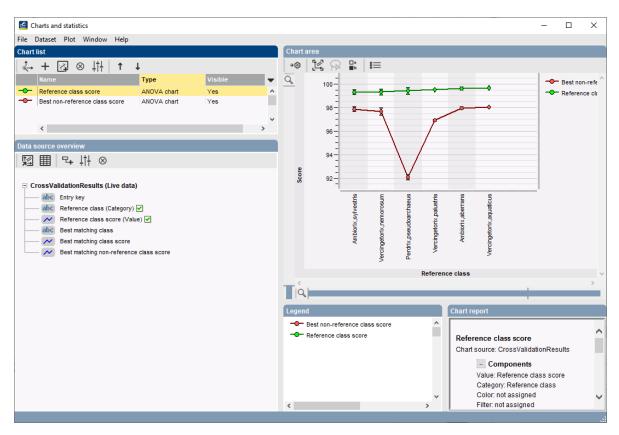


Figure 9: A graph comparing the reference class score (green) with the best non-reference class score (red).

6. Close both the graph and the *Identification cross validation* window, save the identification project (*File* > *Save* (B, Ctrl+S)) and close it. We are now ready to identify unknown samples.

6 Identifying unknown samples

In the database, there are several entries that have not been identified down to species level, for these entries the field **Species** contains **sp.**.

Before running the identification, we will create a field to contain the results.

- 1. Highlight the *Database entries* panel, select *Edit* > *Information fields* > *Add information field...* and type in **ID results** as name. Leave all other settings at default and click <*OK* >.
- Make sure no entries are selected using *Database* > *Entries* > *Unselect all entries (all levels)* (F4).
- 3. Select all entries with **sp.** for the field **Species** by selecting *Edit* > *Find object in list...* (**E**, **Ctrl+Shift+F**) and typing in **sp.**. Click <*Select all*>.

There should now be five entries selected in the Database entries panel.

- 4. Start the identification wizard using *Analysis* > *Identify selected entries...* (19).
- 5. Make sure the option *Stored classifier* is checked in the first step and press < *Next* > twice.

The *Identification* window will open with the results of the identification (see Figure 10).

dentification						-)
View Window	Help						
	-						
intries		_			Results		
Key	Modified date	Genus	Species	Strain numbe 🔻	Balanced similarity on 16S rDNA (Simi		
G@Gel07@006	2009-06-19 12h29m41s -60	Ambiorix	sp.	52415	[Ambiorix,sylvestris] 96.6		
G@Gel07@008	2009-06-19 12h29m41s -60	Ambiorix	sp.	52424	[Ambiorix,sylvestris] 96.8		
G@Gel08@008	2009-06-19 12h29m41s -60	Perdrix	sp.	25693	[Ambiorix,sylvestris] 93.6		
G@Gel08@012	2009-06-19 12h29m41s -60	Perdrix	sp.	53175	[Perdrix,pseudoarchaeus] 93.1		
G@Gel11@002	2009-06-19 12h29m41s -60	Ambiorix	sp.	52440	[Ambiorix,sylvestris] 96.9		
<pre>etails for G@Gel1*</pre>	l@002 / Balanced similarity	on 16S rDNA	_	, ,	Settings		
	l@002 / Balanced similarity	on 16S rDNA	-	>	Settings		
etails for G@Gel1'	@002 / Balanced similarity	r on 16S rDNA		> •	Reference comparison: RefSet		
etails for G@Gel1'		r on 16S rDNA			Reference comparison: RefSet Reference labels: Comparison groups		
letails for G@Gel1 lass mbiorix,sylvestris mbiorix,aberrans	Similarity 96.9 96.3	r on 16S rDNA		•	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity		
etails for G@Gel1 ass mbiorix,sylvestris mbiorix,aberrans erdrix,pseudoarchae	Similarity 96.9 96.3 15 91.9	r on 16S rDNA	_	•	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 16S rDNA		
etails for G@Gel1 ass mbiorix,sylvestris mbiorix, aberrans erdrix, pseudoarchae ercingetorix, palustris	Similarity 96.9 96.3 91.9 82.2	r on 16S rDNA		•	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 16S rDNA Data set: Similarity matrix		
etails for G@Gel11 Ress Ilass Imbiorix, sylvestris mbiorix, aberrans erdrix, pseudoarchaei ercingetorix, nemoros ercingetorix, nemoros	Similarity 96.9 96.3 91.9 82.2 um 82.0	r on 16S rDNA	-	•	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 16S rDNA		
etails for G@Gel1 blass mbiorix,sylvestris mbiorix,aberrans erdrix,pseudoarchaer ercingetorix,palustris ercingetorix,nemoros	Similarity 96.9 96.3 91.9 82.2 um 82.0	r on 16S rDNA		÷	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 16S rDNA Data set: Similarity matrix		
etails for G@Gel1 ass mbiorix,sylvestris mbiorix, aberrans erdrix,pseudoarchaet ercingetorix,palustris ercingetorix,nemoros	Similarity 96.9 96.3 91.9 82.2 um 82.0	r on 16\$ rDNA		•	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 165 rDNA Data set: Similarity matrix Similarity coefficient Standard		
etails for G@Gel11	Similarity 96.9 96.3 91.9 82.2 um 82.0	on 16S rDNA		÷	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 165 rDNA Data set: Similarity matrix Similarity coefficient Standard Pairwise alignment		
etails for G@Gel11	Similarity 96.9 96.3 91.9 82.2 um 82.0	r on 16S rDNA		÷	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 165 rDNA Data set: Similarity matrix Similarity coefficient Standard		
etails for G@Gel11 lass mbiortx,sylvestris mbiortx,aberrans rdrix, paeudoarchae ercingetorix, palutris ercingetorix, nemoros ercingetorix, aquaticu latch comments	Similarity 96.9 96.3 96.3 91.9 22.2 00.0			÷	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 163 rDNA Data set: Similarity matrix Similarity coefficient Standard Pairwise alignment Open gap penalty: 100 % Unit gap penalty: 00 %		
etails for G@Gel1 Elass mbiork, sylvestrs mbiork, aberrans rdrix, pseudoarchaele ercingefork, palustris ercingefork, nemoros ercingefork, aquaticu atch comments "Similarity" score i	Similarity 96.9 96.3 91.9 92.2 91.9 82.2 91.9 82.2 93.9			• •	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Balanced similarity Experiment type: 165 rDNA Data set: Similarity matrix Similarity coefficient Standard Pairwise alignment Open gap penalty: 100 % Unit gap penalty: 00 % Similarity calculation		
etails for G@Gel1 Elass mbiork, sylvestrs mbiork, aberrans rdrix, pseudoarchaele ercingefork, palustris ercingefork, nemoros ercingefork, aquaticu atch comments "Similarity" score i	Similarity 96.9 96.3 96.3 91.9 22.2 00.0		96	• •	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Elatanced similarity Experiment type: 16S rDNA Data set Similarity coefficient Standard Pairwise alignment Open gap penalty: 100 % Unit gap penalty: 0 % Similarity calculation Cost table: Default		
etails for G@Gel1 Eas mbiork, sylvestris mbiork, aberrans erdrk, pseudoarchaer erdrk, pseudoarchaer erdrk, pseudoarchaer erdringefork, nemoros erdingefork, nemoros erdingefork, nemoros erdingefork, aquaticu latch comments "Similarity" score i "Similarity" score of	Similarity 96.9 96.3 91.9 92.2 91.9 82.2 91.9 82.2 93.9	atches is below 1	96	• •	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Elanced similarity Experiment type: 16S rDNA Data set Similarity matrix Similarity coefficient Standard Pairwise alignment Open gap penalty: 100 % Unit gap penalty: 00 % Similarity calculation Cost table: Default Use conversion cost only: No		
etails for G@Gel1 lass mbiork, sylvestris mbiork, aberrans erdrix, paeudoarchaer erdringetork, nemoros ercingetork, nemoros ercingetork, aquaticu tatch comments "Similarity" score i "Similarity" score o — Compa	Similarity 96.9 96.3 96.3 91.9 82.2 82.2 82.0 81.2	atches is below 1	96	• •	Reference comparison: RefSet Reference labels: Comparison groups Classifier type: Elatanced similarity Experiment type: 16S rDNA Data set Similarity coefficient Standard Pairwise alignment Open gap penalty: 100 % Unit gap penalty: 0 % Similarity calculation Cost table: Default		

Figure 10: The identification results.

In this case, it is clear that based on the 16s rDNA sequences, the unknown samples cannot be assigned to any species as they do not meet the similarity threshold. The highest hit is shown in the right panel, but is in grey between brackets (see Figure 10).

For Figure 11 two samples that had already been identified were added to illustrate the difference between an assigned hit and no assignment.

When transferring identification results to the database, results between brackets will be transferred as <**Undetermined**>. More detailed results on the highlighted entry are shown in the

Identification							-	
: View Window H 副 卣 │↑+ ↓-	elp							
intries					Results			
Key	Modified date	Genus	Species	Strain nun 🔫	Balanced similarity on 16S	rDNA		
G@Gel07@006	2009-06-19 12h29m41s -60	Ambiorix	sp.	52415	[Ambiorix,sylvestris]	96.6		
G@Gel07@008	2009-06-19 12h29m41s -60	Ambiorix	sp.	52424	[Ambiorix,sylvestris]	96.8		
G@Gel08@003	2009-06-19 12h29m41s -60	Perdrix	pseudoarchaeus	25671	Perdrix,pseudoarchaeus	99.9		
G@Gel08@008	2009-06-19 12h29m41s -60	Perdrix	sp.	25693	[Ambiorix,sylvestris]	93.6		
G@Gel08@012	2009-06-19 12h29m41s -60	Perdrix	sp.	53175	[Perdrix,pseudoarchaeus]	93.1		
G@Gel09@003	2009-06-19 12h29m41s -60	Perdrix	pseudoarchaeus	25676	Perdrix,pseudoarchaeus	99.9		
G@Gel11@002	2009-06-19 12h29m41s -60	Ambiorix	sp.	52440	[Ambiorix,sylvestris]	96.9		
<				>				
etails for G@Gel08@0	003 / Balanced similarity on 16	S rDNA			Settings			
照								
Class	Similarity			-	Reference comparison: Re			
Perdrix,pseudoarchaeus	99.9			^	Reference labels: Compar			
Ambiorix, aberrans	92.0				Classifier type: Balanced s			
	91.7				Experiment type: 16S rDNA			
Ambiorix,sylvestris	01.1							
/ercingetorix,nemorosum	83.4				Data set: Similarity matrix			
/ercingetorix,nemorosum /ercingetorix,palustris	83.4 82.2			-		fficient		
/ercingetorix,nemorosum /ercingetorix,palustris	83.4				Data set: Similarity matrix	fficient		
/ercingetorix,nemorosum /ercingetorix,palustris	83.4 82.2			v	Data set: Similarity matrix - Similarity coe Standard			
/ercingetorix,nemorosum /ercingetorix,palustris /ercingetorix,aquaticus	83.4 82.2	_		~	Data set. Similarity matrix Similarity coe Standard Pairwise	e alignment		
Ambiorix, sylvestris Vercingetorix, nemorosum Vercingetorix, palustris Vercingetorix, aquaticus Match comments	83.4 82.2		_	~	Data set: Similarity matrix Standard Open gap pe	e alignment nalty: 100 %		
/ercingetorix,nemorosum /ercingetorix,palustris /ercingetorix,aquaticus	83.4 82.2	_	_	~	Data set. Similarity matrix Similarity coe Standard Pairwise	e alignment nalty: 100 %		
/ercingetorix, nemorosum /ercingetorix, palustris /ercingetorix, aquaticus Aatch comments	83.4 82.2 82.1	sults		~	Data set: Similarity matrix Similarity coe Standard Pairwis Open gap pe Unit gap pen	e alignment nalty: 100 %		
/ercingetorix, nemorosum /ercingetorix, palustris /ercingetorix, aquaticus Match comments	83.4 82.2 82.1	sults	-	-	Data set: Similarity matrix Similarity coe Standard Pairwis Open gap pe Unit gap pen	e alignment nalty: 100 % atty: 0 % ty calculation		
/ercingetorix, nemorosum /ercingetorix, palustris /ercingetorix, aquaticus Match comments — Comparis Average score:	83.4 82.2 82.1 on with cross-validation re 99.411085 ± 0.30081	sults		v	Data set: Similarity matrix Similarity coe Standard Pairwis Open gap pe Unit gap pen Similari Cost table. D	e alignment nalty: 100 % atty: 0 % ty calculation		
Vercingetorix, nemorosum Vercingetorix, palustris Vercingetorix, aquaticus Match comments — Comparis Average score:	83.4 82.2 82.1	sults	-	v	Data set: Similarity matrix Similarity coe Standard Pairwis Open gap pe Unit gap pen Similari Cost table. D	e alignment naity: 100 % alty: 0 % ty calculation efault on cost only: No		
Vercingetorix, nemorosum Vercingetorix, palustris Vercingetorix, aquaticus Match comments — Comparis Average score:	83.4 82.2 82.1 on with cross-validation re 99.411085 ± 0.30081	sults		v	Data set Similarity matrix Standard Pairwise Open gap pe Unit gap pen Similarit Cost table: D Use convers Correction: N	e alignment naity: 100 % alty: 0 % ty calculation efault on cost only: No		

Figure 11: Identification results.

center panel. A comparison with this entry and all members of a certain group in the reference set can also be opened from this panel, this allows the user to evaluate the results in more detail.

6. To open a comparison with a certain entry and the members of a certain group, highlight this entry, select a group in the panel with the detailed results (center panel) and select View > Open in comparison window ().

A comparison window opens with all members of the highlighted group and the highlighted entry.

The results of the identification can be transferred to the database using *File* > *Transfer results* to database (B), select the field ID results and click < OK >.