



ISSN: 1608-9391
e-ISSN: 2664-2786

Received:16/12/2020
Accepted:13/2/2021

Specialized High-Quality 16S Ribosomal RNA Gene Databases for Identification of Bacterial Taxonomic Groups: a review

*Talal S. Salih

**Hiba K. Saeed

Department of Biophysics/ College of Science/ University of Mosul

***Mohammed A. Ibraheem

Department of Biology/ College of Science/ University of Mosul

*E-mail: talal.salih@uomosul.edu.iq

**E-mail: hiba.khaleel@uomosul.edu.iq

***E-mail: mohrsbio122@uomosul.edu.iq

ABSTRACT

Recent accumulation of the bacterial 16S and 23S rRNA gene sequences in the public GenBank databases obtained from the wide use of the Polymerase Chain Reaction (PCR) and the Next Generation Sequencing (NGS) applications has led the specialists in the fields of microbiology and bioinformatics to provide specific and reliable databases for storing, retrieving and aligning bacterial ribosomal gene sequences. In our review, we have provided a comprehensive description of gene databases with an emphasis on the most specialized and high quality rRNA gene databases. These databases are freely accessible and widely used for all specialist users for accurate bacterial genera and species identification from different sources. rRNA genes, platforms, internet addresses and references of each gene specialized database were outlined. The Silva, EzTaxon-e and BIBI have been shown to be the most confident and reliable specialized rRNA gene databases. These databases contain the semi fully high-quality ribosomal gene sequences along with some house-keeping and uncultured gene sequences in case of EzTaxon-e.

Keywords: Bacterial Identification, GenBank Databases, Protein Coding Genes, rRNA Genes, Specialized Gene Databases.

BACKGROUND

The 16S ribosomal RNA (rRNA) gene has become the gold standard and the reliable method for detection and identification of bacterial genera and in some cases specific-species (Petti, 2007; Tringe and Hugenholtz, 2008; Vinje *et al.*, 2015). The bacterial 16S rRNA molecule is the most conserved marker gene that shows considerable sequence diversity among different bacteria (Chakravorty, 2007). There are a number of reasons that favour the extensive utilization of the 16S rRNA to detect and identify bacterial genera and species. These reasons include first, it is universally distributed in bacteria, thus the lineage among all bacteria can be measured (Woese *et al.*, 1985); second, it is conserved over time, which enabled the design of suitable primers for PCR amplification; third, its size (~1,500bp) is suitable for sequencing and provides enough information for analysis (Patel, 2001; Mizrahi-Man *et al.*, 2013). In 1982, only 606 sequences were available which were deposited in GenBank, the largest global database of gene sequencing, while as of August 2020 over one and a half billion have been deposited (<https://www.ncbi.nlm.nih.gov/genbank/statistics/>) of which more than 6,000,000 are nucleotide sequences of the 16S rRNA gene. In the last two decades, tremendous amount of sequence data including 16S rRNA genes have been accumulated in the public gene databases as results of the advent of next generation sequencing (NGS) methodologies that targeted amplicon sequencing of the 16S rRNA gene (Reuter *et al.*, 2015; Myer *et al.*, 2016; Rausch *et al.*, 2019; Brumfield *et al.*, 2020; Winand *et al.*, 2020). Furthermore, megabases of 16S rRNA gene sequence libraries from the environmental samples have been cloned and deposited in the GenBank databases (Sczyrba *et al.*, 2017; Escobar-Zepeda *et al.*, 2018; Reza *et al.*, 2018; Bharti and Grimm, 2019; Chen *et al.*, 2020). In this work, we provide a brief review on the specialized high-quality and freely available 16S rRNA sequence gene databases, such as Silva database, EzTaxon-e database, BIBI database and other similar specialized databases. This review summarizes their basic function, unique features and retrieving sequences utility. The genes, platforms, URLs and references of all databases discussed in this review can be found in (Table 1).

Table 1: Overview, internet addresses and references of 16S rRNA sequence gene databases

Database	URL	Reference	No. of citation
Silva	https://www.arb-silva.de/	Quast <i>et al.</i> , 2013; Yilmaz <i>et al.</i> , 2014	10451
EzTaxon-e	https://www.ezbiocloud.net/	Chun <i>et al.</i> , 2007; Yoon <i>et al.</i> , 2017	5141
BIBI	https://umr5558-bibiserv.univ-lyon1.fr/lebibi/lebibi.cgi	Devulder <i>et al.</i> , 2003; Flandrois <i>et al.</i> , 2015	218
RDP	https://rdp.cme.msu.edu/	Cole <i>et al.</i> , 2009; Cole <i>et al.</i> , 2014	7275
Greengenes	http://greengenes.lbl.gov/	DeSantis <i>et al.</i> , 2006	4372

Importance of the Specialized 16S Gene Databases

Due to the rapid increase of 16S gene sequences in the last few years that are deposited in public databases, it was no longer be manually inspected and processed. This laid the foundation of specialized 16S gene databases that are able to mimic the manual curation process and align millions of ribosomal gene sequences. The 16S gene sequences in these databases are checked for quality, taxonomy and annotations (Glöckner *et al.*, 2017). These databases are designed to cover the 16S gene sequence of all the domain of life with a coherent and an easily accessible way. Access to the ribosomal gene sequences in all of these databases is free for all users via the web (Santamaria *et al.*, 2012). These databases contain useful information about the 16S gene sequences such as the validity sequence names and sample from which it obtained, and the gene variants. Most of the public databases contain a number of partial and poorly annotated 16S gene sequences

(Edgar, 2018). On the other hand, the specialized 16S gene databases contain complete or almost complete ribosomal gene sequences (Pruesse *et al.*, 2012; Flandrois *et al.*, 2015). This often employs the discriminatory power to identify bacteria at the genus or the species level. Finally, these gene databases contain sequences from different clinical and environmental sources including cultured and uncultured species obtained from different sequence platforms including next generation sequencing and metagenomics.

Silva Database

Silva database (Quast *et al.*, 2013), is a comprehensive biological resource that actively incorporates (16S/18S) and (23S/28S) high-quality ribosomal RNA gene sequences for all domains of life (Bacteria, Archaea and Eukaryota). This is a unique feature among the current existing rRNA specialized databases which contain only rRNA sequences. The Silva was used for quality checked and aligned sequenced regions of both 16S and 23S rRNA. Silva database contained, when it was released in 2007, 353,366 16S and 46,979 23S rRNA sequences (Pruesse *et al.*, 2007). The current Silva database release 138 (December 2019) contains 9, 469, 656 and 907, 382 rRNA gene sequences. The datasets for each gene in Silva are either Parc or Ref. The Silva Parc dataset comprises the full-length sequences for the respective genes; the Silva Ref comprises nearly full-length high-quality sequences for the respective gene (Glöckner *et al.*, 2017). The input data file should be in FASTA format, whereas the corresponding output datasets can be downloaded as ARB, zip, tgz as well as FASTA format. The Silva database can be readily used with the standalone program or as a web-based service. The ribosomal RNA gene sequence data in SILVA are retrieved from the EMBL/ENA nucleotide sequence database (Kanz *et al.*, 2005). The SILVA database is developed and maintained by the Microbial Genomics and Bioinformatics Research Group in Bremen, Germany, in cooperation with the Department of Microbiology at the Technical University Munich and the Ribocon GmbH.

EzTaxon-e Database

EzTaxon-e (Chun *et al.*, 2007; Yoon *et al.*, 2017) is a specialized 16S rRNA gene sequences database for the identification of bacteria and archaea type strains. The database is based on pairwise nucleotide similarity values and phylogenetic inference methods for the identification of isolates. While the original version of EzTaxon database contained information on type strains, the last version of the database provides an extended information by including 16S rRNA gene sequences of unclassified or uncultured species (Kim *et al.*, 2012). This is a unique useful feature for studying the taxonomy and identification of cultured and uncultured species. The current EzTaxon-e database contains 2, 943, 209 16S rRNA gene sequences of the bacteria and archaea domains. EzTaxon-e is a web-based service and cannot be used standalone. The input data file can be either in FASTA or in ABI format, whereas the corresponding output datasets can be downloaded as Excel, EzEditor2 as well as FASTA format. All sequence data in EzTaxon-e are basically collected from the NCBI GenBank (Benson *et al.*, 2012). Moreover, the database utilizes genome and metagenome sequencing projects for retrieving the 16S rRNA gene sequences using the rRNASelector program (Lee *et al.*, 2011). The EzTaxon database is produced and maintained by ChunLab, Inc., South Korea.

BIBI Database

The Bioinformatics Bacterial Identification database (BIBI) (Devulder *et al.*, 2003) is another specialized biological database that was developed for bacterial identification based on ribosomal RNA gene sequences. The original version of this database was based on 16S rRNA sequences which were extracted from the public NCBI GenBank database (Benson *et al.*, 2012). The new version of BIBI database now uses 16S and 23S rRNA gene sequences as well as protein coding gene sequences of *gyrB*, *recA*, *sodA*, *rpoB*, *tmRNA*, *tuf* and *groel2-hsp65* (Flandrois *et al.*, 2015). Besides the NCBI GenBank database, the new gene sequences of subsets that are involved in the new version of BIBI database are retrieved from the EMBL/ENA nucleotide sequence database

(Kanz *et al.*, 2005) and from the RNAcentral, a new source of information for RNA gene sequences, a comprehensive database (Gorodkin and Seemann, 2019). BIBI database uses a series of nucleotide sequences in FASTA format as an input data file. A unique feature of BIBI system is that the output results are displayed in a tree in NEWICK format after retrieving the most closely related strain sequences to the query sequence, aligning them, and performing phylogenetic reconstruction using an approximate maximum likelihood method (Flandrois *et al.*, 2015). The BIBI database can be facilely performed as a web-based service for a single sequence or batches of sequences submission. The BIBI database is developed and maintained by the Bacteriology Lab, Faculty of Medicine, University of Claude Bernard – Lyon, France.

Similar Specialized Databases

Two of the most worldwide similar specialized ribosomal RNA gene Databases are the Ribosomal Database Project (RDP) database (Cole *et al.*, 2014) and the Greengenes database (DeSantis *et al.*, 2006).

The RDP database contains 16S rRNA gene sequences for bacteria and archaea, as well as 28S gene sequences for Fungi. The database lacks sequences of 23S for bacteria and archaea and 18S for Fungi. The gene sequences of RDP database are coming from cultured and uncultured microorganisms. Most of rRNA gene sequences in the RDP are retrieved from the NCBI GenBank database (Benson *et al.*, 2012), the EMBL/ENA nucleotide sequence database (Kanz *et al.*, 2005), the DDBJ database (Mashima *et al.*, 2016) and from the direct submissions to the RDP system (Maidak *et al.*, 2000). The last RDP database release 11.5 (September 2016) contains 3,356,809 aligned and annotated bacterial and archaeal 16S rRNA sequences and 125,525 Fungal 28S rRNA sequences. The input data file can be in FASTA, GenBank or in EMBL format, whereas the corresponding output datasets are sequences in FASTA format. The RDP system database can be used with the standalone programs and as a web-based reference. The RDP database is hosted and maintained by the Center for Microbial Ecology at Michigan State University, USA.

On the other hand, the Greengenes database provides 16S rRNA gene sequences for bacteria and archaea domains (DeSantis *et al.*, 2006). The database lacks any 23S rRNA or eukaryotic gene sequences. The rRNA gene sequences in Greengenes database are obtained from the NCBI GenBank database (Benson *et al.*, 2012) and CyanoDB database (Caporaso *et al.*, 2010). The Greengenes database can readily be used with the standalone program and as a web-based service. It is hosted and maintained by the Lawrence Berkeley National Laboratory, USA. However, the database seems to be no longer being maintained, it has not been updated for the last eight years.

Table 2: Gene types, platforms and hosts of the specialized 16S rRNA gene databases

Database	Gene	Platform	Host
Silva	rRNA and rRNA	Standalone program and web-based service	Microbial Genomics and Bioinformatics Research Group in Bremen and Department of Microbiology at the Technical University Munich and the Ribocon GmbH / Germany
EzTaxon-e	rRNA	Web-based service	ChunLab, Inc., South Korea
BIBI	16S rRNA , 23S rRNA and coding gene sequences of <i>gyrB</i> , <i>recA</i> , <i>sodA</i> , <i>rpoB</i> , <i>tmRNA</i> , <i>tuf</i> , <i>groel2-hsp65</i>	Web-based service	The Bacteriology Lab, Faculty of Medicine, University of Claude Bernard – Lyon, France.
RDP	16S rRNA and 23S rRNA	Standalone program and web-based service	The Center for Microbial Ecology at Michigan State University, USA
Greengenes	16S rRNA	Standalone program and web-based service	The Lawrence Berkeley National Laboratory, USA

CONCLUSION

In our review we have provided a detailed description of the most robust and comprehensive rRNA gene sequences databases that are used to retrieve, align and compare the query rRNA genes for bacterial identification. We summarized type of rRNA genes, platforms, internet addresses and references of each gene specialized database. It was very obvious that Silva database superior over the rest databases as it includes the full-length sequences for all domains of life (Bacteria, Archaea and Eukaryota). However, other databases such as EzTaxon-e database with a unique feature of identification of uncultured novel species based on 16S rRNA obtained from NGS and metagenomics methods and the BIBI database with a unique feature of constructing a maximum likelihood phylogenetic tree besides the bacterial species identification have preferable utilization with those who are interested in isolation and identification of novel uncultured bacterial species or those who are less familiar of using advanced bioinformatics tools for drawing and constructing the phylogenetic trees, respectively.

REFERENCES

- Benson, D.A.; Karsch-Mizrachi, I.; Clark, K.; Lipman, D.J.; Ostell, J.; Sayers, E.W. (2012). GenBank. *Nucleic Acids Res.*, **40**, D48–53.
- Bharti, R.; Grimm, D.G. (2019). Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics.*, **12**, 1-16.
- Brumfield, K. D.; Huq, A.; Colwell, R.R.; Olds, J.L.; Leddy, M. B. (2020). Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *Plos. one.*, **15**, e0228899.
- Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Pena, A.G.; Goodrich, J.K.; Gordon, J.I.; Huttley, G.A. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods.*, **7**, 335.
- Chakravorty, S.; Helb, D.; Burday, M.; Connell, A.; Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiolog. Methods.*, **6**, 330–339
- Chen, Q.; Meyer, W.A.; Zhang, Q.; White, J.F. (2020). 16S rRNA metagenomic analysis of the bacterial community associated with turf grass seeds from low moisture and high moisture climates. *Peer J.*, **8**, e8417.
- Chun, J.; Lee, J.H.; Jung, Y.; Kim, M.; Kim, S.; Kim, B.K.; Lim, Y.W. (2007). EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *International J. Systematic and Evolution. Microbiol.*, **57**, 2259-2261.
- Cole, J. R.; Wang, Q.; Cardenas, E.; Fish, J.; Chai, B.; Farris, R.J.; Tiedje, J. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucl. Acids Research.*, **37**, D141-D145.
- Cole, J.R.; Wang, Q.; Fish, J.A.; Chai, B.; McGarrell, D.M.; Sun, Y.; Brown, C.T.; Porras-Alfaro, A.; Kuske, C.R.; Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucl. Acids Research.*, **42**, D633-D642.
- DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D., Hu, P.; Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, **72**, 5069-5072.
- Devulder, G.; Perriere, G.; Baty, F.; Flandrois, J.P. (2003). BIBI, a bioinformatics bacterial identification tool. *J. Clin. Microbiol.*, **41**, 1785-1787.
- Edgar, R. C. (2018) Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *Peer J.*, **6**, e4652
- Escobar-Zepeda, A.; Godoy-Lozano, E. E.; Raggi, L.; Segovia, L.; Merino, E.; Gutiérrez-Rios, R. M.; Sanchez-Flores, A. (2018). Analysis of sequencing strategies and tools for taxonomic annotation: defining standards for progressive metagenomics. *Scient. Reports.*, **8**, 1-13.

- Flandrois, J.P.; Perrière, G.; Gouy, M. (2015). leBIBI QBPP: a set of databases and a webtool for automatic phylogenetic analysis of prokaryotic sequences. *BMC bioinform.*, **16**, 251.
- Glöckner, O.; Yilmaz, P.; Quast, C.; Gerken, J.; Beccati, A.; Ciuprina, A.; Bruns, G.; Yarza, P.; Peplies, J.; Westram, R.; Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.*, **261**, 169-176.
- Gorodkin, J.; Seemann, E.S. (2019). RNAcentral: a hub of information for non-coding RNA sequences. *Nucl. Acids Research.*, **47**, D1-D13.
- Kanz, C.; Aldebert, P.; Althorpe, N.; Baker, W.; Baldwin, A.; Bates, K.; Browne, P.; van den Broek, A.; Castro, M.; Cochrane, G.; Duggan, K. (2005). The EMBL nucleotide sequence database. *Nucl. Acids Research*, **33**, D29-D33.
- Kim, O.S.; Cho, Y.J.; Lee, K.; Yoon, S.H.; Kim, M.; Na, H.; Park, S.C.; Jeon, Y.S.; Lee, J.H.; Yi, H.; Won, S. (2012). Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *International J. System. and Evolution. Microbiol.*, **62**, 716-721.
- Lee, J.H.; Yi, H.; Chun, J. (2011). rRNASelector: a computer program for selecting ribosomal RNA encoding sequences from metagenomic and met transcriptomic shotgun libraries. *The J. Microbiol.*, **49**, 689-697.
- Maidak, B.L.; Cole, J.R.; Lilburn, T.G.; Parker J.T.; Saxman, P.R.; Stredwick, J.M.; Garrity, G.M.; Li, B.; Olsen, G.J.; Pramanik, S.; Schmidt, T.M. (2000). The RDP (ribosomal database project) continues. *Nucl. Acids Research.*, **28**, 173-174.
- Mashima, J.; Kodama, Y.; Kosuge, T.; Fujisawa, T.; Katayama, T.; Nagasaki, H.; Okuda, Y.; Kaminuma, E.; Ogasawara, O.; Okubo, K.; Nakamura, Y. (2016). DNA data bank of Japan (DDBJ) progress report. *Nucl. Acids Research*, **44**, D51-D57.
- Mizrahi-Man, O.; Davenport, E.R.; Gilad, Y. (2013). Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *PLoS ONE.*, **8**, 1-14.
- Myer, P.R.; Kim, M.; Freetly, H.C.; Smith, T.P. (2016). Evaluation of 16S rRNA amplicon sequencing using two next-generation sequencing technologies for phylogenetic analysis of the rumen bacterial community in steers. *J. Microbiology. Methods.*, **127**, 132-140.
- Patel, J.B. (2001). 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *J. Molecul. Diagn.*, **6**, 313-321.
- Petti, C.A. (2007). Detection and identification of microorganisms by gene amplification and sequencing. *Clin. Infect. Dis.*, **44**, 1108-1114.
- Pruesse, E.; Peplies, J.; Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinform.*, **28**, 1823-1829.
- Pruesse, E.; Quast, C.; Knittel, K.; Fuchs, B.M.; Ludwig, W.; Peplies, J.; Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl. Acids Research.*, **35**, 7188-7196.
- Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Research.*, **41**, D590-D596.
- Rausch, P.; Rühlemann, M.; Hermes, B.M.; Doms, S.; Dagan, T.; Dierking, K.; Heinsen, F.A. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome.*, **7**, 1-19.
- Reuter, J.A.; Spacek, D.V.; Snyder, M.P. (2015). High-throughput sequencing technologies. *Molecul. cell.*, **58**, 586-597.
- Reza, M.S.; Mizusawa, N.; Kumano, A.; Oikawa, C.; Ouchi, D.; Kobiyama, A.; Sato, S. (2018). Metagenomic analysis using 16S ribosomal RNA genes of a bacterial community in an urban stream, the Tama River, Tokyo. *Fisher. Sci.*, **84**, 563-577.

- Santamaria, M.; Fosso, B.; Consiglio, A.; De Caro, G.; Grillo, G.; Licciulli, F.; Pesole, G. (2012). Reference databases for taxonomic assignment in metagenomics. *Briefings in Bioinform.*, **13**, 682-695.
- Sczyrba, A.; Hofmann, P.; Belmann, P.; Koslicki, D.; Janssen, S.; Dröge, J.; Gregor, I.; Majda, S.; Fiedler, J.; Dahms, E.; Bremges, A. (2017). Critical assessment of metagenome interpretation—a benchmark of metagenomics software. *Nature Methods.*, **14**, 1063-1071.
- Tringe, S. G.; Hugenholtz, P. (2008). A renaissance for the pioneering 16S rRNA gene. *Current Opinion in Microbiol.*, **11**, 442-446.
- Vinje, H.; Liland, K.H.; Almøy, T.; Snipen, L. (2015). Comparing K-mer based methods for improved classification of 16S sequences. *BMC Bioinform.*, **16**, 1-13.
- Winand, R.; Bogaerts, B.; Hoffman, S.; Lefevre, L.; Delvoye, M.; Van Braekel, J.; Vanneste, K. (2020). Targeting the 16s rRNA gene for bacterial identification in complex mixed samples: Comparative evaluation of second (illumina) and third (oxford nanopore technologies) generation sequencing technologies. *International J. Molecul. Sci.*, **21**, 298-312
- Woese, C.; Stackebrandt, E.; Macke, T.; Fox, G. (1985). A phylogenetic definition of the major eubacterial taxa. *Systemic and Appl. Microbiol.*, **6**, 143-151
- Yilmaz, P.; Parfrey, L. W.; Yarza, P.; Gerken, J.; Pruesse, E.; Quast, C.; Glöckner, F. O. (2014). The SILVA and “all-species living tree project (LTP)” taxonomic frameworks. *Nucl. Acids Research.*, **42**, D643-D648.
- Yoon, S.H.; Ha, S.M.; Kwon, S.; Lim, J.; Kim, Y.; Seo, H.; Chun, J. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International J. Systemat and Evolut. Microbiol.*, **67**, 16-23.

قواعد بيانات الجينات الحامض النووي الريبوزي الريبوسومي المتخصصة عالية الجودة 16S RNA لتشخيص الاجناس البكتيرية وتحديد الانواع: مقال

الملخص

أدى التراكم الأخير للتسلسل الجيني البكتيري للحامض النووي الريبوزي الريبوسومي الصغير 16S والكبير rRNA S23 في قواعد بيانات بنك الجينات العامة التي تم الحصول عليها من الاستخدام الواسع لتطبيقات تفاعل البلمرة المتسلسل (PCR) وتطبيقات تسلسل الجيل القادم الى توجيه المتخصصين في مجالات علم الأحياء المجهرية والمعلوماتية الحيوية لتوفير قواعد بيانات محددة وموثوقة لتخزين واسترجاع وتنظيم تسلسل جينات الحامض الريبوزي البكتيري. في هذا المقال: قدمنا وصفاً شاملاً مع التركيز على قواعد بيانات الجينات للحامض النووي الريبوزي الريبوسومي الأكثر تخصصاً وعالية الجودة. قواعد البيانات هذه متاحة مجاناً وتستخدم على نطاق واسع لجميع المستخدمين المتخصصين لتشخيص الاجناس البكتيرية الدقيقة والأنواع من مصادر مختلفة. تم تحديد جينات الحامض الريبوزي الريبوسومي، والمنصات، وعناوين الإنترنت، ومصادر كل قاعدة بيانات متخصصة في الجينات. ثبت أن Silva و EzTaxon-e و BIBI هي قواعد بيانات الجينات المتخصصة الأكثر دقة وموثوقية. تحتوي قواعد البيانات هذه على تسلسل الجينات الريبوزية الريبوسومية شبه عالية الجودة جنباً إلى جنب مع بعض الموروثات الخدمية و التسلسلات الجينية كما في حالة EzTaxon-e.

الكلمات الدالة: تشخيص البكتيريا، قواعد بيانات بنك الجينات، الجينات المشفرة للبروتين، جينات الحامض النووي الريبوزي الريبوسومي، قواعد بيانات الجينات المتخصصة.