

# BIOLOGICAL SEQUENCE & SEQUENCE ALIGNMENT pada Stanieria cyanosphaera

Disusun Oleh:

Nama: Safira Raissa Rahmi

NPM :2006568891



# Daftar Isi

01

DNA Sequencing

02

Biological  
Sequences

03

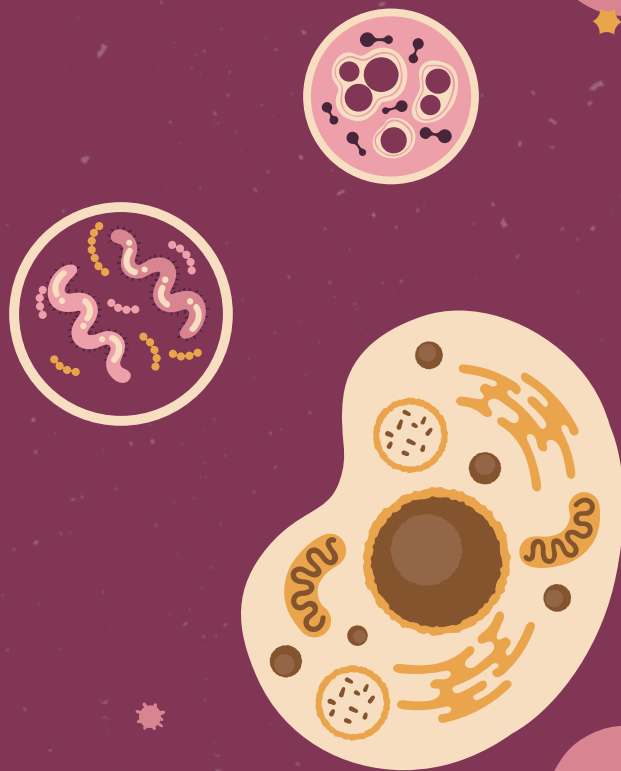
Sequencing  
Alignment

04

BLAST Search  
Tool

01

# DNA Sequencing



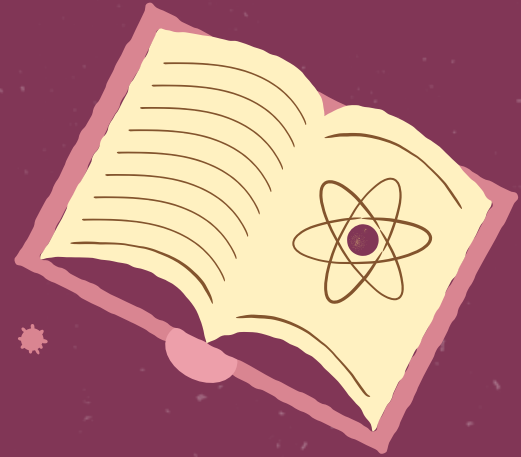
# DNA

- **DEFINISI**

DNA adalah asam nukleat yang mengandung informasi genetik yang diturunkan dari generasi ke generasi dan bersifat individual.

- **FUNGSI**

DNA menyimpan dan mengkodekan instruksi genetik dan menentukan hubungan antar organisme.



# DNA Sequencing

Proses penentuan urutan nukleotida  
(A, T, C, dan G) pada suatu potongan  
DNA.



# Komponen DNA Sequencing

Sampel DNA



Nukleotida



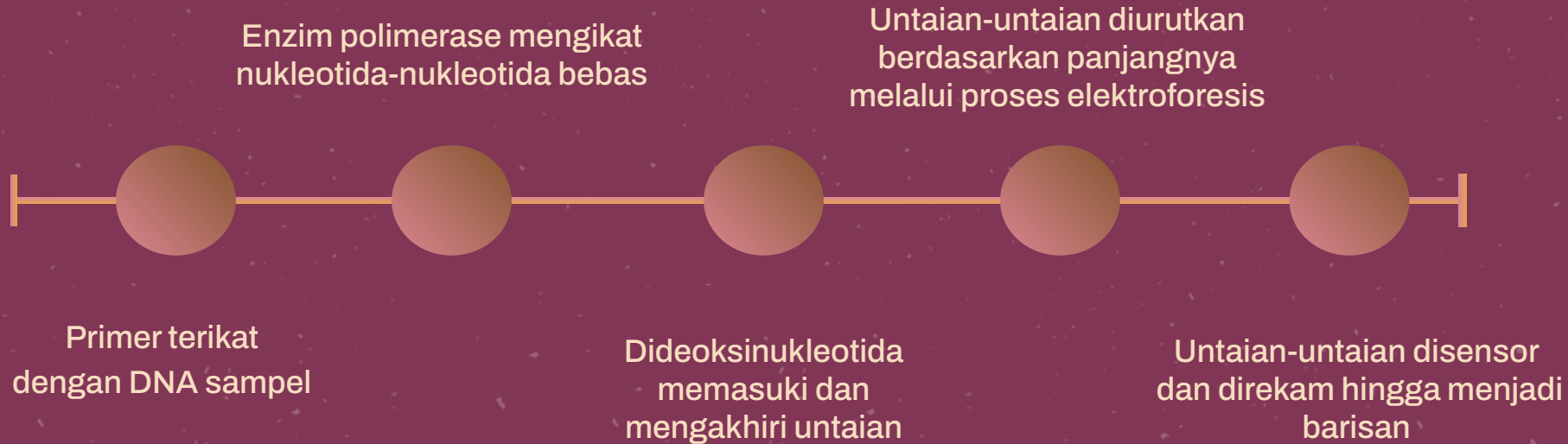
Enzim  
Polimerase



Didieoksi  
Nukleotida



# PROSES DNA SEQUENCING



# Manfaat DNA Sequencing

**Penelitian  
Biologi Dasar**



**Bidang  
Terapan**

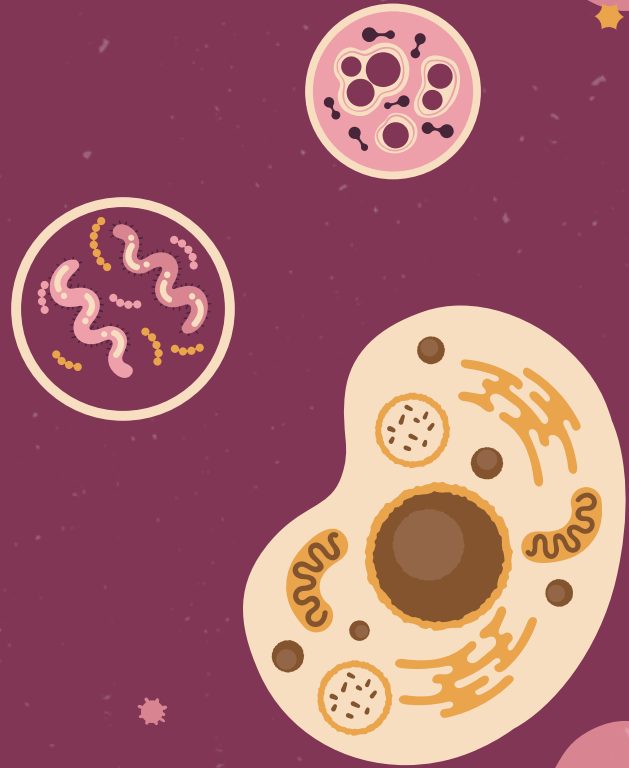


**Genografi  
DNA**



02

# Biological Sequences





“Molekul asam nukleat atau protein **tunggal** yang berkesinambungan yang dipandang sebagai hierarki kelas **dengan** pewarisan **berganda**.”

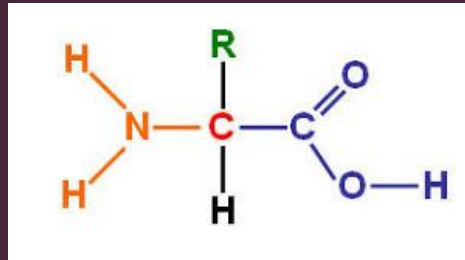
—Definisi

# 3 JENIS BIOLOGICAL SEQUENCES

**RNA**



**Protein**



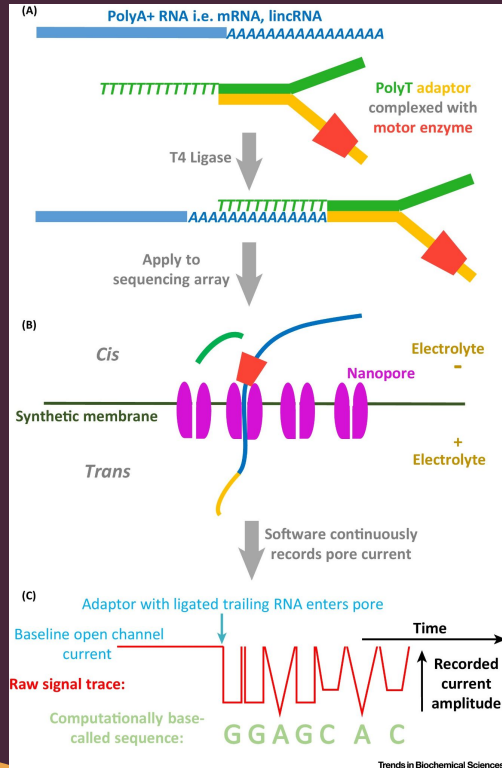
**DNA**





Dalam DNA, terdapat sekuens yang berhubungan dengan alfabet DNA, yaitu A (Adenin), C (Sitosin), G (Guanin), dan T (Timin). Alfabet DNA tersebut adalah asam nukleat atau nukleotida. Penentuan sekuens nukleotida yang terdapat pada DNA disebut DNA sequencing.

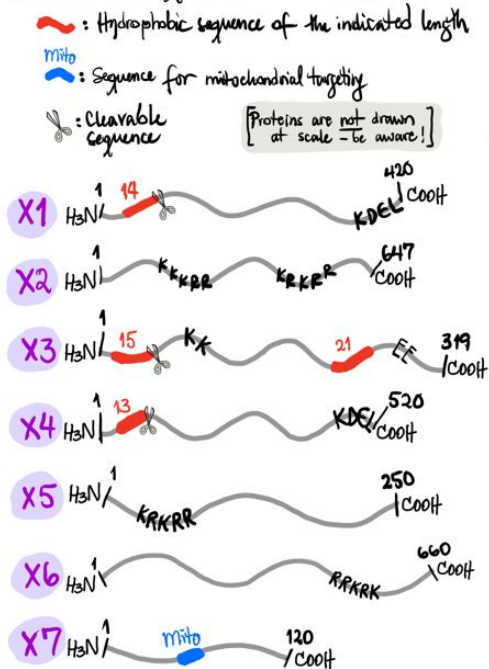
# RNA SEQUENCES



RNA sequencing adalah teknik yang dapat meneliti kuantitas dan sekuens RNA dalam sebuah sampel menggunakan *next-generation sequencing*, berlawanan dengan Sanger sequencing. Kemudian, RNA sequencing memungkinkan kita untuk menyelidiki dan menemukan transkriptom.

# PROTEIN SEQUENCES

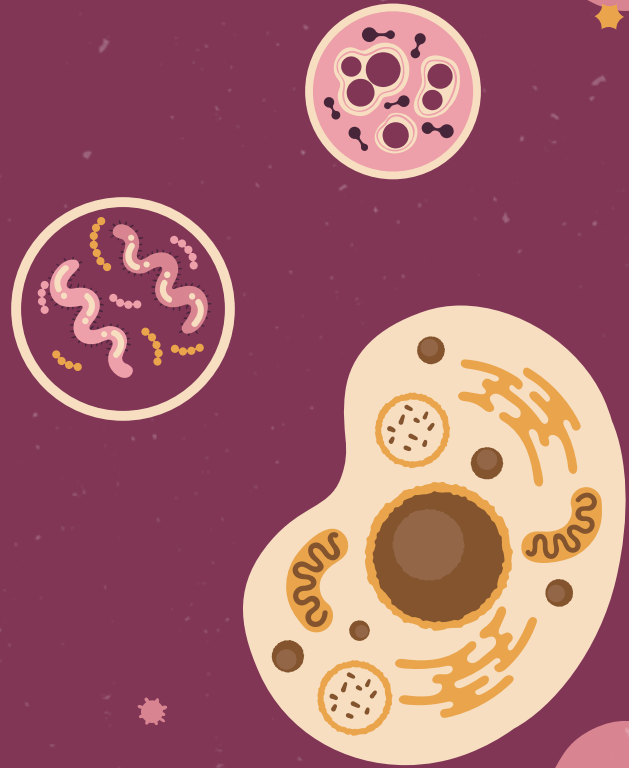
## Protein sequence data:



Setiap protein atau peptida memiliki sekuens linear asam amino. Struktur dari protein berkemungkinan langsung tersekuens atau tersimpul dari sekuens DNA. Proses menentukan sekuens asam amino disebut protein sequencing.

03

# Sequencing Alignment





“Sebuah metode menyelaraskan urutan DNA, RNA, atau protein untuk mengidentifikasi wilayah kesamaan yang mungkin **dihasilkan** dari hubungan fungsional, struktural, atau evolusi antara urutan.”

—Definisi



# Metode Sequencing Alignment



**Local**

Algoritma  
Smith-Waterman



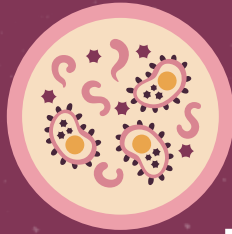
**Global**

Algoritma  
Needleman-Wunsch



✱  
✱ Algoritma Smith-Waterman mencari  
✱ **keselarasan lokal terbaik**, yaitu  
**pencocokan substring dalam dua**  
sekuens, sedangkan algoritma  
Needleman-Wunsch mencari  
**keselarasan global terbaik**, yaitu  
**pencocokan end-to-end** dari sekuens  
yang **terlibat**.

Perbedaan Local dan Global



# LOCAL ALIGNMENT

Algoritma Smith-Waterman merupakan perluasan dari algoritma **text or sequence string matching** sebagai salah satu implementasi program dinamis. Algoritma ini dikembangkan **pada tahun 1981** oleh Temple F. Smith dan Michael S. **Waterman**. Algoritma ini **didasarkan pada** local alignment, yang dilakukan dengan membagi **sequence** menjadi beberapa **sub bagian**.

# Tahapan Local Alignment

## 1. Definisikan dulu:

- a. Pencocokan nilai setiap karakter terjadi ketika ada dua karakter yang sama dengan nilai positif.  
Karena dua teks dianggap lebih mirip jika nilai kecocokannya tinggi.
- b. Nilai ketidakcocokan karakter diatur ke negatif atau nol. Nilai ini juga harus valid secara simetris (karakter a dan b memiliki nilai -10, sehingga konflik antara karakter b dan a juga harus -10)
- c. Nilai penalti dihasilkan ketika salah satu karakter dari dua teks yang akan dibandingkan digeser sehingga diganti dengan tanda baca kosong yang harus bernilai negatif.

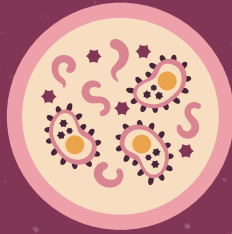
# Tahapan Local Alignment

2. Setelah ketiga nilai ini diberikan, mereka membentuk matriks dengan jumlah baris = panjang teks pertama + 1 dan jumlah kolom = panjang teks kedua + 1.

$$F(i, j) = \max \begin{cases} 0, \\ F(i-1, j-1) + s(x_i, y_j), \\ F(i-1, j) - d, \\ F(i, j-1) - d. \end{cases}$$

Saat mengisi matriks, salah satu nilai matriks tidak boleh negatif, sehingga diasumsikan 0 bisa menjadi nilai maksimum.

Karena tidak ada nilai yang kurang dari nol, ini memungkinkan algoritma Smith-Waterman untuk fokus hanya pada wilayah dengan urutan yang serupa. Nilai positif diberi nilai 1, jadi nilai dalam array hanyalah bilangan biner.



# GLOBAL ALIGNMENT

Algoritma Needleman-Wunsch merupakan perluasan dari algoritma **text or line order matching** sebagai salah satu implementasi program **dinamik**. Algoritma ini dikembangkan oleh Saul B. Needleman dan Christian D. Wunsch pada tahun 1970. Algoritma ini digunakan untuk mencari **keselarasan global terbaik**, yaitu **mencocokkan panjang urutan** dari ujung ke ujung.

# Tahapan Global Alignment

## 1. Definisikan dulu:

- a. Nilai untuk setiap kecocokan karakter: harus bilangan positif. Skor pertandingan karakter bisa sama untuk setiap karakter.
- b. Nilai ketidakcocokan karakter  $s(x_i, y_i)$ : negatif atau nol. Nilainya negatif karena mengurangi kemiripan teks. Nilai misfit simetris.
- c. Nilai penalti (d) jika salah satu karakter dari dua teks yang dibandingkan dipindahkan sedemikian rupa sehingga diganti dengan tanda baca kosong: harus negatif. Ini karena kita membutuhkan usaha ekstra untuk memindahkan karakter setelah celah yang ditambahkan.

## 2. Diberikan dua urutan $x_i = x_1, x_2, \dots, x_n$ , dengan panjang $n$ , dan urutan $y_j = y_1 y_2 \dots y_m$ dengan format

Matriks panjang  $m$  yang ukurannya  $(n+1)$  kali  $(m+1)$ .

# Tahapan Global Alignment

3. Nilai pada elemen kolom ke-j dan baris ke-i dari matriks ditentukan oleh fungsi berikut.

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - d \\ F(i, j-1) - d \end{cases}$$

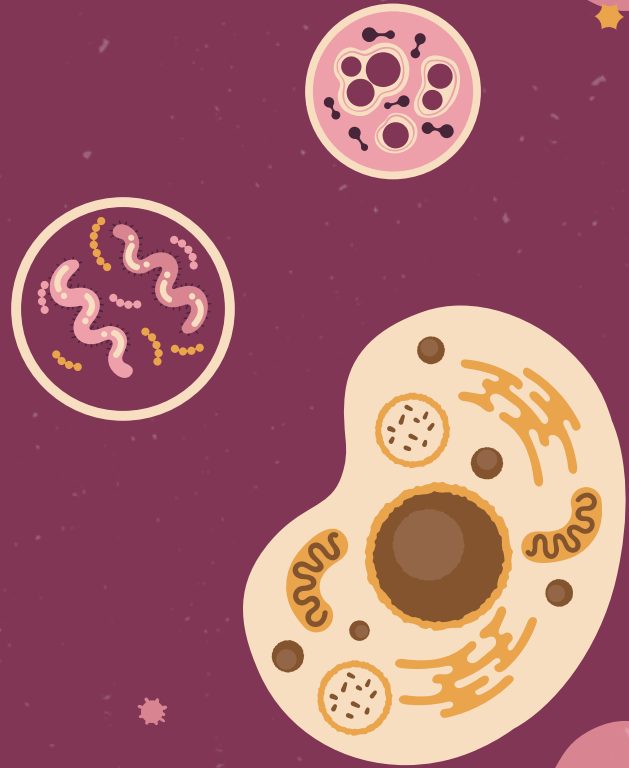
Setelah elemen matriks diisi, nilai elemen matriks baris (n+1) dan kolom (m+1) menunjukkan nilai korespondensi terbesar.

4. Telusuri lintasan dari elemen baris - n+1 dan kolom - m+1 (misalnya  $F_{ij}$ ) ke elemen baris 0 dan kolom 0 (indeks baris dan kolom mulai dari 0).



04

# BLAST Search Tools



# Query

TGGCGGACGGGTGAGTAAAGCGTGAGAATCTGCTCTAGGTCGGGGACAACAGTTGAAACGACTGCTAATCC  
CGGATGAGCCTGTGGGTAAAAGATTAATTGCCTAGAGAGGAGCTCGCGTCTGATTAGCTAGTTGGTGAGGTAAA  
GGCTCACCAAGGCGACGATCAGTAGCTGTTCTGAGAGGAAGAACAGCCACACTGGGACTGAGACACGGCCCA  
GACTCTACGGGAGGCAGCAGTGGGGAAATTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGAG  
GGAAGAAGGCCTTTGGGTCGTAAACCTCTTTTCTCAGGGAAGAAGTTCTGACGGTACCTGAGGAAAAAGCCTC  
GGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGAGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCG  
TCCGTAGGTGGCCTTTCAAGTCTGCGGTTAAAAACTAGAGCTCAACTCTGGGTCAGCCGTGGAAACTGAGAAG  
CTAGAGTACAGTAGGGGTAGAGGGAATTCCAGTGTAGCGGTGAAATGCGTAGAGATTGGGAAGAACACCGGT  
GGCGAAAGCGCTCTACTGGGCTTGTACTGACACTGAGGGACGAAAGCTAGGGGAGCAAAAGGGATTAGATAC  
CCCTGTAGTCCTAGCCGTAAACGATGGAAGTGGCGTAGCCTGTATCAACTCGGGCTGTGCCGAAGCTAACGC  
GTTAAGTTTCCCGCTGGGGAGTACGCACGCAAGTGTGAAACTCAAAGGAATTGACGGGGGCCCCGACAAGC  
GGTGGAGTATGTGGTTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCGCGAATCTCAGGGA  
AACTTGAGAGTGCCTAAGGGAACGCGGAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCTGAGATGTTG  
GGTTAAGTCCCGCAACGAGCGCAACCTCGTCCTTAGTTGCCATCATTAAAGTTGGGAAGTCTAGGGAGACTGCC  
GGGGACAAGTCCGAGGAAGGTGGGGATGACGTCAAGTCAGCATGCCCTTACGTCTGGGCTACACACGTACT  
ACAATGGTTGAGACAAAAGGCGAGCAACTCGCGAGAGCCAGCGAATCCAGCAAAGTACAGCCCCAGTTCAGA  
TTGCAGGCTGCAACTCGCTGCATGAAGTAGGAATCGCTAGTAATCGCCGGTCAGCATACGGCGGTGAATCCGT  
TCCCGGGCCTTGTACACACCGCCCGTCACACCATGGAAGCTGGAACATCCGAAGTCGTTACTCCAACCCGCAA  
GGGAGGAGGGCGCCGAAGGTGGGGCTAGTGACTGGGGTGAAGTCGTAACAAGGTAGCCGTACCGGAAGGTG  
TGGCTGGATCACCTCCTTAAAGGGAGACCTGTTGAATGGAGAGAAACCAATAGCCTCCATGAAAACACAAGC  
AGGTACAGATTGGGGGACTAAACTTTCAAACATTAGATGGGGTTACCTATCAAAGAAAACGGGTATTAGCTC  
AGGTGGTTAGAGCGCACCCCTGATAAGGGTGAGGTCCCTGGTTCGAGTCCAGGATGGCCACCACCCCGAAAG  
AGACCCAGCAATTGAGGAACTCAACCTGCTGGATTCTCGAATCCAGTCCAGAACCTTGAAAAGTGCATAGAAGA  
AAAAAGAGCCACTCAAAATAGAGAAAAGCTGGAATCAATCGAAAGAGAGAAGAAAGCGAAAAACTATGAGTAG  
GGAGCTCACAGACAATC

# HASIL BLAST RELATED SEQUENCE



☒ select all 100 sequences selected

[GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA View](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">Stanieria cyanosphaera PCC 7437, complete genome</a>	<a href="#">Stanieria cyano...</a>	2032	4061	90%	0.0	89.65%	5041209	<a href="#">CP003653.1</a>
<input checked="" type="checkbox"/>	<a href="#">Spirulina sp. PCC 6313 partial ribosomal RNA operon, strain PCC 6313</a>	<a href="#">Spirulina major...</a>	2012	2148	84%	0.0	92.33%	4617	<a href="#">AM709631.1</a>
<input checked="" type="checkbox"/>	<a href="#">Spirulina sp. genes for 16S rRNA, tRNA-Ile, 23S rRNA</a>	<a href="#">Spirulina major...</a>	1991	2127	84%	0.0	92.12%	1970	<a href="#">X75045.1</a>
<input checked="" type="checkbox"/>	<a href="#">Spirulina major strain PCC 6313 16S ribosomal RNA, partial sequence</a>	<a href="#">Spirulina major</a>	1977	1977	79%	0.0	92.12%	1457	<a href="#">NR_125712.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cyanothece sp. PCC 8801, complete genome</a>	<a href="#">Rippakea orient...</a>	1947	4170	84%	0.0	91.49%	4679413	<a href="#">CP001287.1</a>
<input checked="" type="checkbox"/>	<a href="#">Stanieria cyanosphaera PCC 7437 16S ribosomal RNA, partial sequence</a>	<a href="#">Stanieria cyano...</a>	1938	1938	79%	0.0	91.67%	1481	<a href="#">NR_102468.1</a>
<input checked="" type="checkbox"/>	<a href="#">Microcystis aeruginosa NIES-843 DNA, complete genome</a>	<a href="#">Microcystis aer...</a>	1932	3854	90%	0.0	88.56%	5842795	<a href="#">AP009552.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cryptochroococcus tibeticus TP201716.4 clone 06 16S ribosomal RNA gene, partial sequence; 16S-23S...</a>	<a href="#">Cryptochroococ...</a>	1917	1917	90%	0.0	88.59%	1767	<a href="#">MN215485.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cyanobacterium aponinum PCC 10605 chromosome, complete genome</a>	<a href="#">Cyanobacteriu...</a>	1917	6106	84%	0.0	91.20%	4114099	<a href="#">CP003947.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cryptochroococcus tibeticus TP201716.4 clone 05 16S ribosomal RNA gene, partial sequence; 16S-23S...</a>	<a href="#">Cryptochroococ...</a>	1914	1914	90%	0.0	88.54%	1766	<a href="#">MN215484.1</a>
<input checked="" type="checkbox"/>	<a href="#">Limnospira fusiformis SAG 85.79 chromosome, complete genome</a>	<a href="#">Limnospira fusif...</a>	1910	3821	90%	0.0	88.44%	6423694	<a href="#">CP051185.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cryptochroococcus tibeticus TP201716.4 clone 02 16S ribosomal RNA gene, partial sequence; 16S-23S...</a>	<a href="#">Cryptochroococ...</a>	1910	1910	90%	0.0	88.53%	1766	<a href="#">MN215483.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arthrospira sp. str. PCC 8005 chromosome, complete genome</a>	<a href="#">Limnospira indi...</a>	1910	3821	90%	0.0	88.44%	6228153	<a href="#">FO818640.1</a>
<input checked="" type="checkbox"/>	<a href="#">Coleofasciculus chthonoplastes SAG 2209 16S ribosomal RNA gene, partial sequence; 16S-23S riboso...</a>	<a href="#">Coleofasciculus...</a>	1910	1910	90%	0.0	88.34%	2022	<a href="#">EF654055.1</a>
<input checked="" type="checkbox"/>	<a href="#">Arthrospira sp. genes for 16S rRNA, tRNA-Ile, tRNA-Ala, 23S rRNA</a>	<a href="#">Arthrospira plat...</a>	1906	1906	90%	0.0	88.37%	1962	<a href="#">X75044.2</a>

# Top Hit BLAST



```
rRNA      complement(207...>1606)
           /locus_tag="Sta7437_R0009"
           /product="16S ribosomal RNA"

ORIGIN
1  ggtgggcat  cctggacttg  aaccagggac  ctcaccctta  tcaggsgtgc  gctctaacca
61  cctgagctaa  tagcccatgt  tctttcaacc  gaaccagtta  ctagtttgaa  ggctttaatt
121  ttttccctag  tccgaccttg  ggattatttg  aatctctact  tatttgtgta  tctttcagta
181  gacttcaaca  ggtctcccta  aaaaggaggt  gatccagcca  cactttccgg  tacggctacc
241  ttgttacgac  ttcaccccag  tcactagccc  tgccttcggc  atccccctcc  ctacgggttt
301  agggtaacga  cttcggcgct  ggcacagctc  catggtgtga  cgggcgggtg  gtacaaggcc
361  cgggaacgga  ttcaccgcag  tatgctgacc  tgcgattact  agcgattctc  ctttcatgca
421  ggcgagttgc  agcctgcaat  ctgaactgag  gcagggtttg  ctgagattcg  cttactcttg
481  cgagcttgct  gccctttgtc  cctaccattg  tagtacgtgt  gtagcccgag  acgtaagggg
541  catgctgact  tgacgtctac  cccacccttc  tccggtttgt  caccggcgag  ctttctagag
601  tgcccagcct  aactgtcggc  aactaagaac  gaggggttgc  ctggttgcgg  gacttaaccc
661  aacatctcac  gacacgagct  gacgacagcc  atgcaccacc  tgtgttcgcg  ctcccgaagg
721  cactctctag  tttcctagtg  attcgcgaca  tgtcaagttc  tgtaaggtt  cttcgcgttg
781  catcgaatta  aaccacatac  tccaccgctt  gtgcggggcc  cgtcaattc  ctttgagttt
841  cacacttgcg  tgcgtactcc  ccaggcggga  tacttaacgc  gttagctcgc  gcaactgctg
901  ggtcgtatca  agcaacgcct  agtatccatc  gtttacagct  aggactacag  gggtatctaa
961  tccctttcgc  tcccctagct  ttgctccctg  agcgtcagtt  acggcccagt  agagcgctt
1021  cgccactggt  gttcttccca  atctctacgc  atttcaccgc  tacactggga  attcctctta
1081  cccctaccgt  actctagttt  cacagtttcc  accgcctgcc  cggagttaag  cttcagcttt
1141  tgacagcaga  cttgtgaaac  gccttcgga  cgctttacgc  ccaataatc  cggataacgc
1201  ttgcatcctc  cgtattaccg  cggtcgtcgg  caccggagta  gccgatgctg  attcctcaag
1261  taccgtcaga  acttcttctc  tgagaaaaga  ggtttacaac  ccaaaggcct  tctccctca
1321  cgcggtattg  ctccgtcagg  ctttcgcca  ttgcggaaaa  ttccccactg  ctgctctccg
1381  taggagttct  ggccgtgtct  cagtcaccgt  gtggctgctc  atctctctag  accagctacc
1441  gatcgtcgcc  ttggtgcgct  ctaccacac  caactagcta  atcggacgcg  agctctcttt
1501  caggcgaatt  acttttacct  ttggcacat  cgggtattag  cagtgtttc  caactgttat
1561  cccgtcctg  aaggcagatt  ctcacgcgtt  atccaccggt  cgcca
```

## Stanieria cyanosphaera PCC 7437, complete genome

GenBank: CP003653.1

[FASTA](#) [Graphics](#)

[Go to:](#) ☐

LOCUS	CP003653	1606 bp	DNA	linear	BCT 22-JUL-2013
DEFINITION	Stanieria cyanosphaera PCC 7437, complete genome.				
ACCESSION	CP003653	REGION: 451313..452918			
VERSION	CP003653.1				
DBLINK	BioProject: <a href="#">PRJNA158877</a> BioSample: <a href="#">SAMN02261352</a>				
KEYWORDS	GSC:MIGS:2.1.				
SOURCE	Stanieria cyanosphaera PCC 7437				
ORGANISM	Stanieria cyanosphaera PCC 7437 Bacteria; Cyanobacteria; Pleurocapsales; Dermocarpellaceae; Stanieria.				
REFERENCE	1 (bases 1 to 1606)				
AUTHORS	Gugger,M., Coursin,T., Rippka,R., Tandeau De Marsac,N., Huntemann,M., Wei,C.-L., Han,J., Detter,J.C., Han,C., Tapia,R., Davenport,K., Daligault,H., Erkkila,T., Gu,W., Munk,A.C.C., Teshima,H., Xu,Y., Chain,P., Chen,A., Krypides,N., Mavromatis,K., Markowitz,V., Szeto,E., Ivanova,N., Mikhailova,N., Ovchinnikova,G., Pagani,I., Pati,A., Goodwin,L., Peters,L., Pitluck,S., Woyke,T. and Kerfeld,C.				
CONSRMT	US DOE Joint Genome Institute				
TITLE	Finished chromosome of genome of Stanieria cyanosphaera PCC 7437				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1606)				

# Hasil Identifikasi

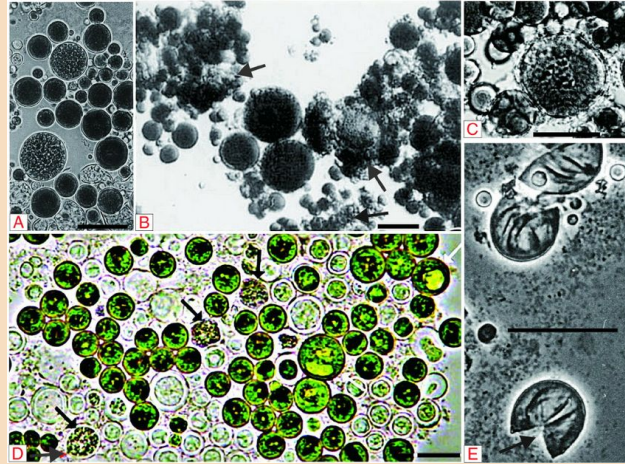


**Search Set:** Database: standard database; Nucleotide collection (nr/nt); Limit to: sequence from type material; Exclude: Models (XM/XP) & Uncultured/environmental sample sequences.

## **BLAST Search Result:**

- Query length: 1770 nt
- Gen 16S rRNA
- Top Hit: *Stanieria cyanosphaera* PCC 7437, complete genome dengan persentase identitas 89,65%
- Max Score: 2032; Total Score: 4061; Query Cover: 90%; E value: 0.0
- Organisme: Bacteria; Cyanobacteria; Pleurocapsales; Dermocarpellaceae; *Stanieria*.

# KARAKTERISTIK



*Stanieria cyanosphaera* adalah baeocyte cyanobacterium, yang bereproduksi melalui beberapa pembelahan dalam tiga bidang untuk membuat sel anak yang lebih kecil. Sel nya berbentuk bulat dengan berbagai ukuran. Diisolasi pertama kali dari kolam di kebun raya, Havana, Kuba pada tahun 1965 dan telah diusulkan sebagai jenis strain.

# DAFTAR PUSTAKA

- Budiman, Ilham Candra. 2009. *Penyejajaran Lokal Barisan DNA dengan Menggunakan Metode Smith Waterman*. Skripsi. Universitas Indonesia : Depok. Diakses dari <http://lib.ui.ac.id/file?file=digital/20181961-009-09-Penyejajaran%20lokal.pdf>.
- Dewi, Alfiana Ririsati. 2018. *Penerapan Algoritma Needleman-Wunsch untuk Mengidentifikasi Mutasi pada Sekuen DNA Virus DNA*. Skripsi. Institut Teknologi Sepuluh Nopember Surabaya. Surabaya.
- Ernawati, Diyah Puspitaningrum, dan Ambar Pravitasari. 2014. *Implementasi Algoritma Smith-Waterman pada Local Alignment dalam Pensejajaran Barisan DNA (Studi Kasus: DNA Tumor Wilms)*. Jurnal Pseudocode, 1, 170-177.
- Muradi, Hengki. 2014. *Matematika Mengungkap Identitas Manusia*. 2014. Diakses dari <https://www.zenius.net/blog/matematika-dna-biologi-identitas-manusia>.
- Zen, Muhamad Reza Firdaus, Sila Wiyanti Putri, dan Muhamad Fajrin Rasyid. 2006. *Penerapan Algoritma Needleman-Wunsch sebagai Salah Satu Implementasi Program Dinamis pada Pensejajaran DNA dan Protein*. Soliton dan DNA. 2006. Diakses pada 29 Mei 2021 dari <http://lipi.go.id/berita/soliton-dan-dna/1135>.
- Needleman-Wunsch Algorithm*. (Last edited 2021, March 24). Wikipedia. [https://en.wikipedia.org/wiki/Needleman%E2%80%93Wunsch\\_algorithm](https://en.wikipedia.org/wiki/Needleman%E2%80%93Wunsch_algorithm)
- Smith-Waterman Algorithm*. (Last edited 2021, May 16). Wikipedia. [https://en.wikipedia.org/wiki/Smith%E2%80%93Waterman\\_algorithm](https://en.wikipedia.org/wiki/Smith%E2%80%93Waterman_algorithm)

Terima  
Kasih!

