

Applying Bioinformatics: The Mysterious Park Fur

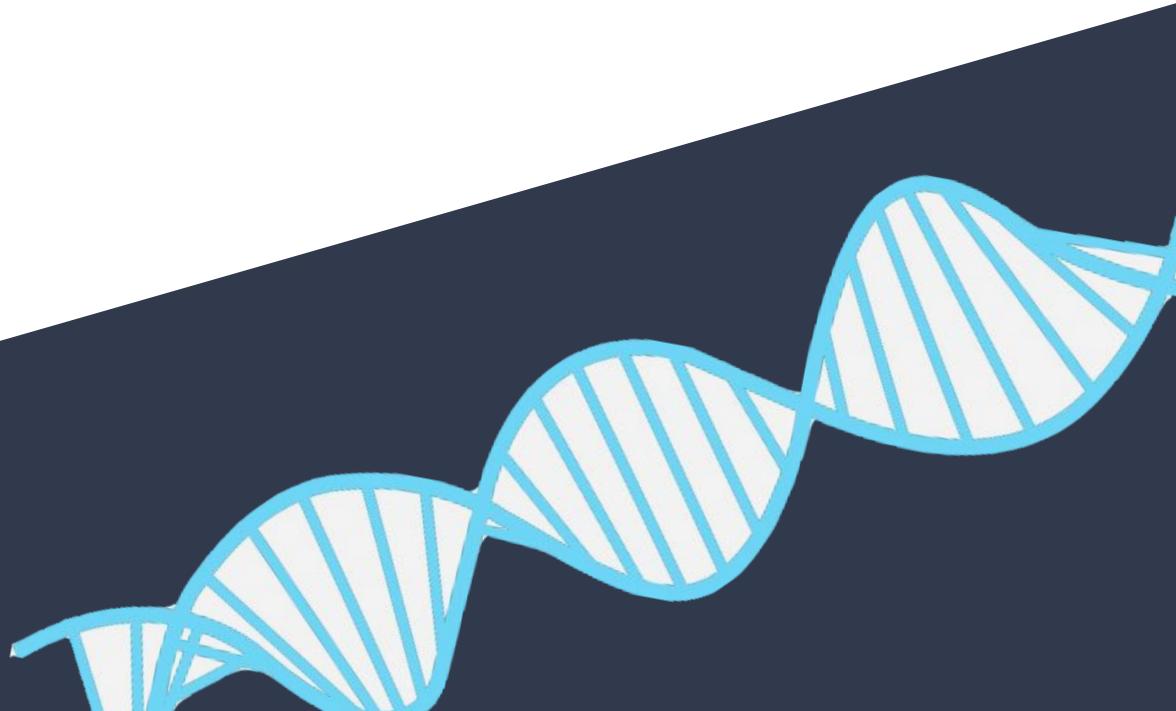
UCSD Graduate School Students:

Jennifer Havens, Sherlyn Weng, Daisy Chen, Lia Gale, Hannah Mummey

Expanding Your Horizons

March 18, 2023

Slides by Caitlin Guccione,
Jennifer Havens, Hannah
Mummey, Kiki Spaulding &
Lauryn Bruce



Who are we?



Job Title: Graduate Student Researchers in
the Bioinformatics Ph.D. program
at University of California, San Diego

What is a researcher?

What is bioinformatics?

What is the path to a Ph.D.?



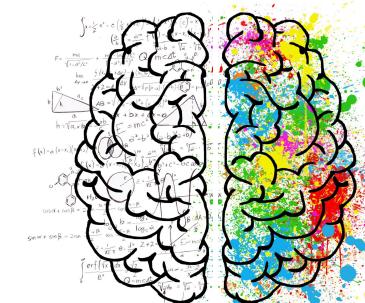
What are components of research?

Social



Mentorship
Teamwork

Creativity



Asking new questions

- Putting together the bigger picture

Analytical
Logic



A1	B	C	D	E
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				

Collecting & Analyzing Data



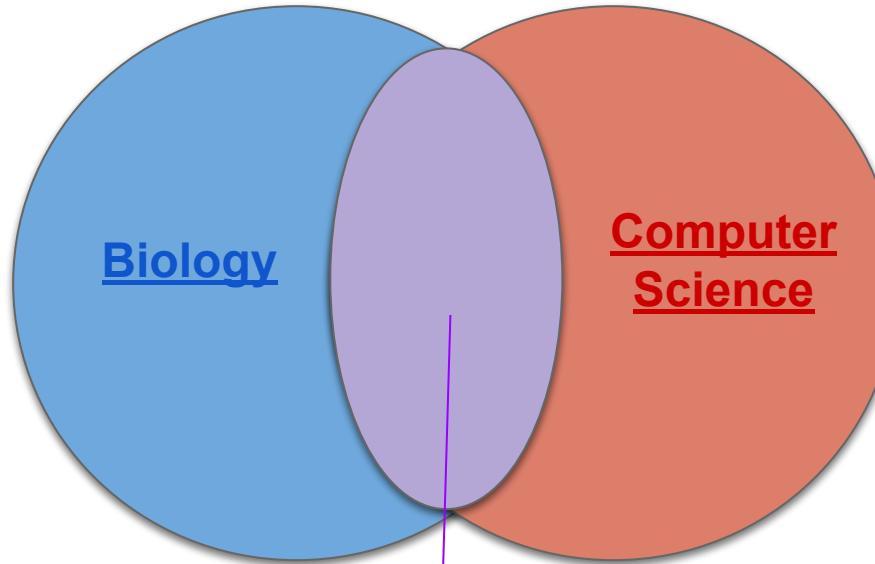
Presenting / Communication / Writing

What is bioinformatics?

How does the body work?

How do diseases work?

How to design new or better therapies?



Handling large datasets

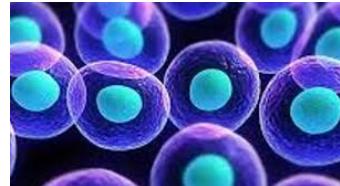
Manipulating data
($+ - \times \div \dots$)

Visualizing data

Answering questions in biology using computer science

What does a bioinformatics researcher do?

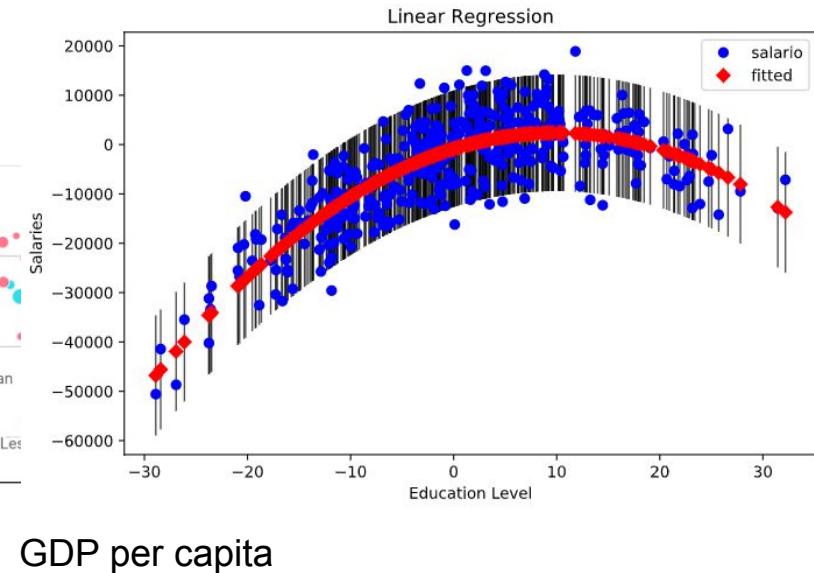
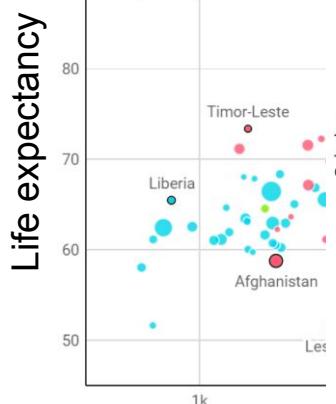
Data collection (Experiments /
Download data)



Data processing & cleaning (Writing code)



Data analysis (Plotting data)



What does a bioinformatics researcher do?

Data interpretation (Understanding data)



Presentation / Publication (Sharing data)



Where you can work with a degree in Bioinformatics?



MD Anderson
~~Cancer Center~~
Making Cancer History®

 Scripps Research

Hospitals



 illumina®



Biotech Companies

 Google Health

 amazon

 fitbit



Tech Companies

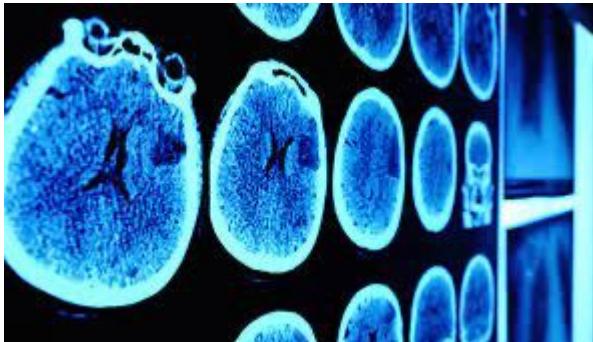
Is working in Bioinformatics fulfilling?



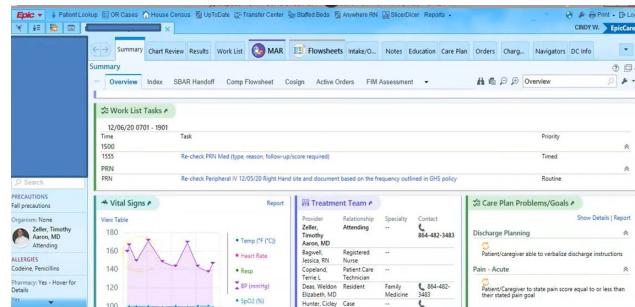
Develop COVID-19 vaccine



Develop new life saving drugs



Help detect disease early to improve outcome



Develop computer programs to help doctors improve patient treatment

NATIONAL CANCER INSTITUTE PRECISION MEDICINE IN CANCER TREATMENT

Discovering unique therapies that treat an individual's cancer based on the specific genetic abnormalities of that person's tumor.

The page features a grid of six rows and three columns of icons. Each icon consists of a stylized human figure, a DNA double helix, and a bottle. The colors of the icons change in a repeating pattern across the grid: blue, purple, green, and orange. The icons represent different aspects of precision medicine: human figures with cancer cells, DNA with a cancer cell, a bottle with a sunburst, and a bottle with a downward-pointing arrow.

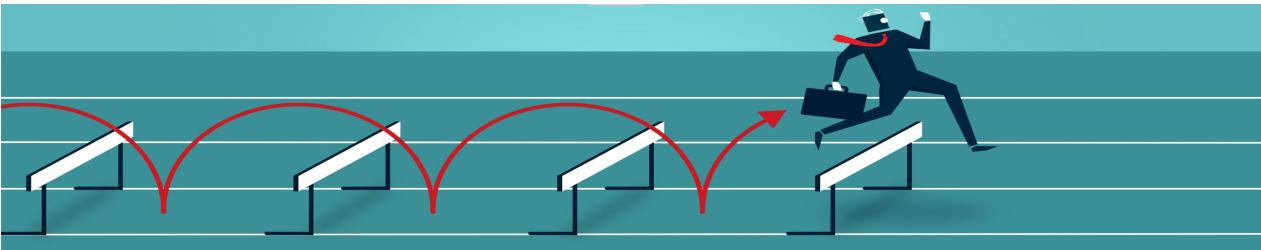
What does the path to a STEM career look like?



School:



Career:



Women in STEM / Bioinformatics

- ~25% of the current bioinformatics workforce is women
- UCSD's BISB 2020 class was the first to have more women than men

Did you ever see yourself in STEM?



Activity

The Mysterious Park Fur

DNA

How many people have heard of DNA?

What does DNA stand for?

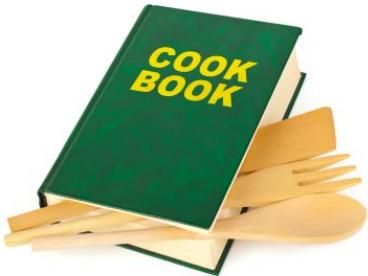
Deoxyribonucleic Acid

What is DNA?

- A molecule inside humans, animals and plants that carries instructions on how to build that organism



What is DNA?



Double-Chocolate Cupcakes Makes 12

Image: ©Ginger Penn-Brown, 2009. Used with permission from allaboutcups.com

TOPPING

1 1/2 cups self-rising flour, sifted	2 1/2 cups confectioners' sugar
1/2 cup unsweetened cocoa powder, sifted	2/3 cup butter, at room temperature
3/4 cup superfine sugar	1 teaspoon vanilla extract
2 large eggs, beaten	Chocolate sprinkles
3/4 cup milk	
1/2 cup butter, melted and cooled	
1 teaspoon vanilla extract	
3/4 cup milk-chocolate chips	

Directions:

- Dredge oven to 400°.
- Line a 12 cup muffin pan with paper cupcake liners.
- Cream flour, cocoa powder and superfine sugar in bowl and make a well in the center.
- In another bowl, beat six eggs, milk, butter and vanilla extract. Add egg mixture to flour mixture. Using a large metal spoon, stir gently to combine.
- Gently fold in chocolate chips.
- Divide batter evenly among cupcake liners and bake until risen and firm to touch, 12–15 minutes. Let cool for a few minutes in pan then transfer to a wire rack.
- Cool completely before frosting.

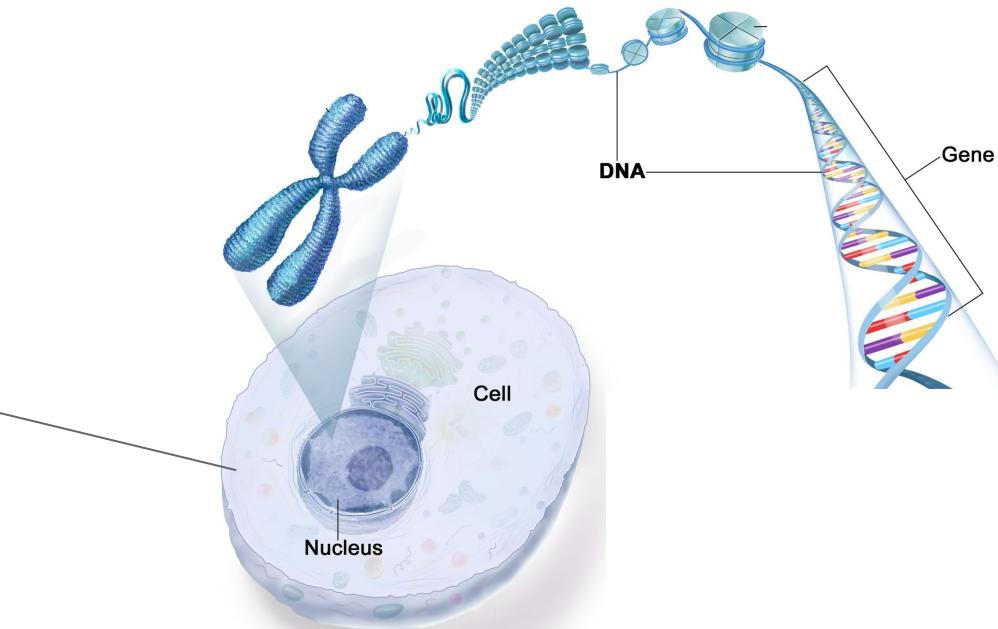
FOR THE TOPPING

- In a bowl combine half each of the confectioners' sugar and butter and mix with wooden spoon until smooth.
- Add remaining confectioners' sugar and butter and beat until light and fluffy.
- Add vanilla extract and mix well.
- Add frosting to a pastry bag fitted with a star tip and pipe onto cupcakes.
- Decorate with chocolate sprinkles.

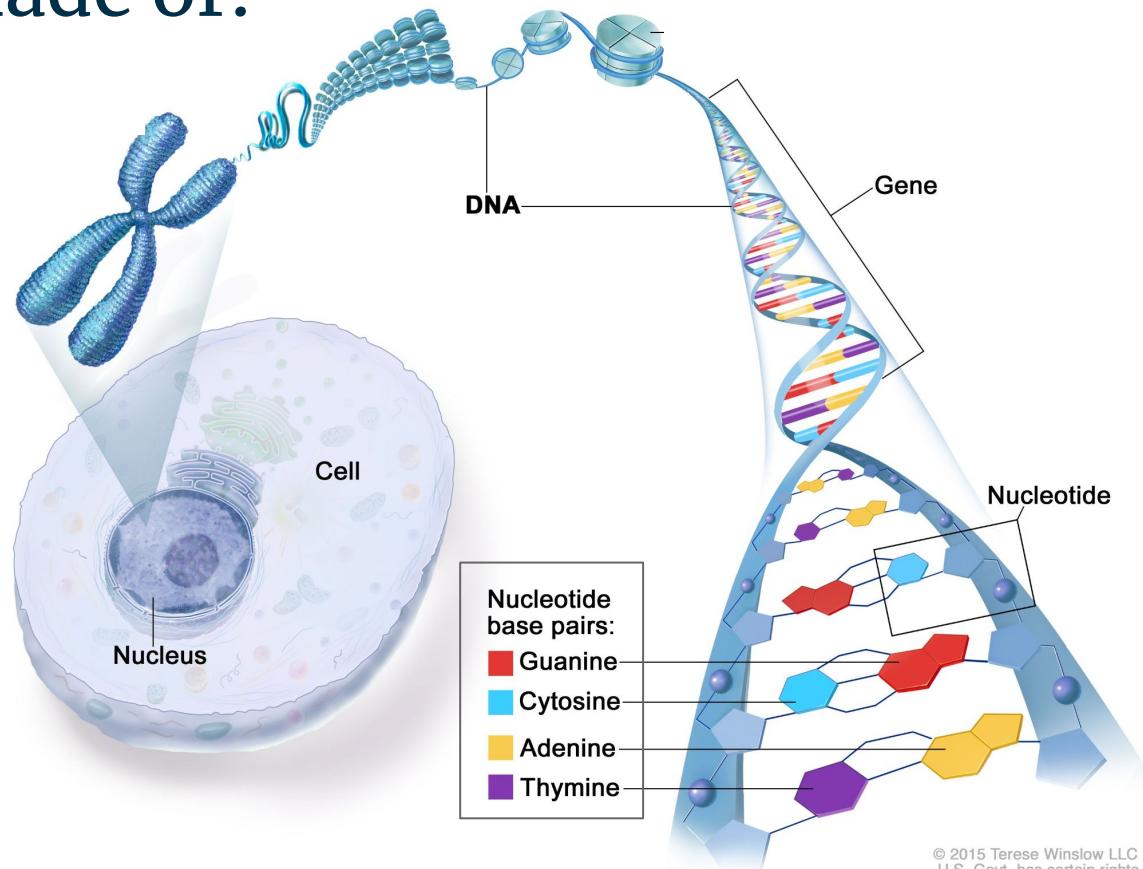
Alicat



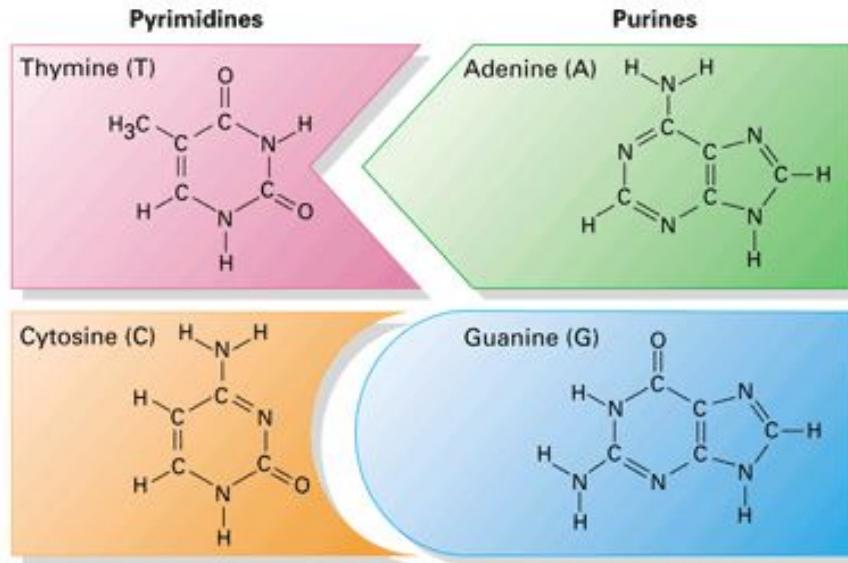
Where is DNA found?



What is DNA made of?



DNA Base Pairs



A C G T T C A
T G C A A G T



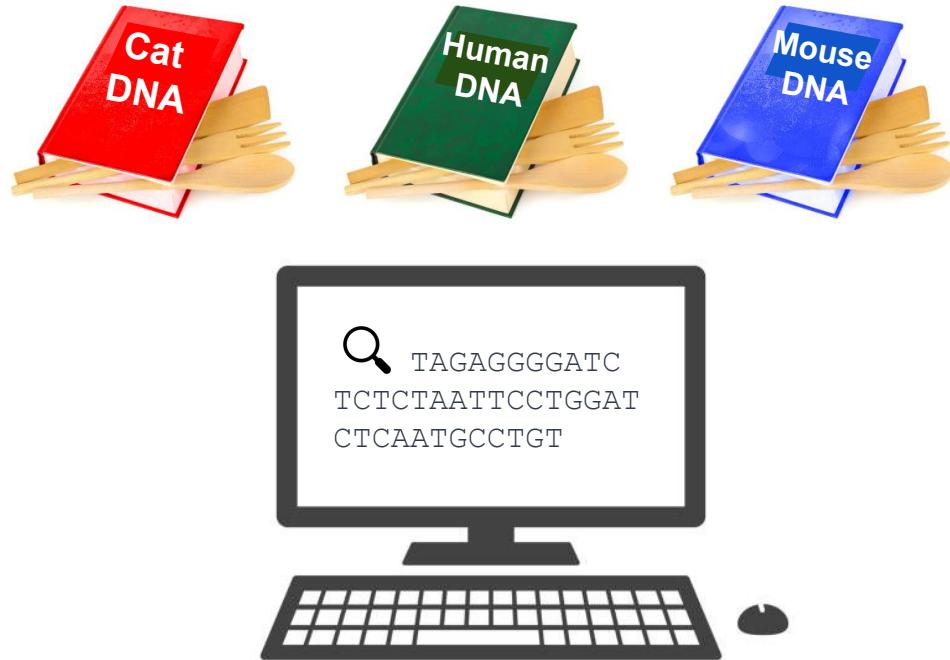
We found some animal fur in the park!
How can we use DNA to figure out what
animal it belongs to?



Match DNA to a database

Unknown DNA sequence

CATTAATATAATATTGTCCCCATATAATTAGAGATT
TCTTCCTCTTGCTGCTTAAGGATCTTCCTTATCT
TTGTAATTGCAAGTTCACTCTAACGTCAAGGTAT
TGAGTGGTTTTATTGACTTAGAGGGGATCTCTCTAA
TTCCTGGATCTCAATGCCTGTTCCCTCCAGGTAAG
GGAAGTTCTCAGCTATGATTTCCTAAATACATATTCT
GGATCTCTGCCCTTCGGCACCCCTCAGGAACCCCAAT
TAAAGGTAGATTTTCCTCTGGCTGTCAATTAT
TTCCCTTAACCTTCCTCATGATCTTTAATTTCCT
CCTTTCTCAGTTCCCTTCCATCAACTTGTCT
TCTATGTCACTCACTGGTTCTCACCTCGTAACCCCT
TGCCTTAGGACCTCTAGTTGGATTGCATCTCATTCA
ATTGATTTTAATTCTGCCTGATTAGATATAATTCTG
CAGTCATGAAGTCTCTGAACCCTTATGCTTTTAT
AGAGCCACCACTAGCTTTAATTGTGCTTCTGAATTG
GCTTCTGACATTGAATTGTAATCAAAGTTGTAACT
CTGTGAGACAGAGGACTGTTCTGATTCTTCTTGAG
GGGAGTTTCCTCTCATCTTCTCAGTGCAGAC
TGGATAAAAACAAG



Compare to DNA sequences from known animals

Using BLAST to match DNA to a database

The screenshot shows the BLAST homepage. At the top left is the NIH logo and "National Library of Medicine" text. To the right are "Log in" and "BLAST®" buttons. Below this is a navigation bar with "Home", "Recent Results", "Saved Strategies", and "Help". The main content area has a title "Basic Local Alignment Search Tool". On the left, a sidebar says "BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance." It includes a "Learn more" link and a "NEWS" section. The news section title is "BLAST+ 2.13.0 is here! Starting with this release, we are including the blastn_vdb and tblastn_vdb executables in the BLAST+ distribution." It also shows the date "Thu, 17 Mar 2022 12:00:00 EST" and a "More BLAST news..." link.

Align: “place or arrange (things) in a straight line”

Database:

AGTTTGATTCA
TTCTTTCGATTGATTCTATTGATTGATTGATTGATTCTTTCTATTCTT
GATTGATTGATTCTT GATTGATTGATTCTT GATTGATTGATTCTT

Your DNA Sequence:

GATTGATTGATTCTT

Mystery fur sequence available at:
https://github.com/jennifer-bio/HWorkshop_mysteryFur

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Copy and paste DNA sequence into search box and press BLAST

BLAST® » blastn suite

Standard Nucleotide BLAST

blastn **blastp** **blastx** **tblastn** **tblastx**

BLASTN programs search nucleotide databases using a nucleotide query.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

```
AGTTTGATTCACTTCTTTGATTGATTCTATTGATTGATTGATTGATTGAT  
TTCTTTCTATTCTTTGATTCAATTGATTCTATTCAATTGATTGAT  
TATTTTGATTAGATTGATTGATTGATTGATTGATTGATTGATTGATTGAT  
ATTCTTTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGAT
```

Query subrange [?](#)

From

To

Or, upload file Choose File No file chosen [?](#)

Job Title Unknown Fur Sample

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

BLAST Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)
 Show results in a new window

+ Algorithm parameters

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Select columns

Show

100

 select all 100 sequences selected[GenBank](#)[Graphics](#)[Distance tree of results](#)[MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Canis lupus genome assembly, chromosome: 15	Canis lupus	1304	24439	100%	0.0	100.00%	65778685	HG994395.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 15a	Canis lupus familiaris	1295	24710	100%	0.0	99.72%	64284575	CP050569.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 15b	Canis lupus familiaris	1295	24675	100%	0.0	99.72%	64292917	CP050620.1
<input checked="" type="checkbox"/>	Canis familiaris, clone XX-388L17, complete sequence	Canis lupus familiaris	983	1204	100%	0.0	91.88%	211225	AC205951.7

Canis lupus genome assembly, chromosome: 15Sequence ID: [HG994395.1](#) Length: 65778685 Number of Matches: 106Range 1: 1680114 to 1680819 [GenBank](#) [Graphics](#)[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1304 bits(706)	0.0	706/706(100%)	0/706(0%)	Plus/Plus

Our unknown sequence

Query 1 CAATAAAGAAGTAAGTATTTAGTATGTCACCCCTGAAAATGCTAACGCATCAAATAAAGC 60

The dog genome

Sbjct 1680114 CAATAAAGAAGTAAGTATTTAGTATGTCACCCCTGAAAATGCTAACGCATCAAATAAAGC 1680173

Query 61 AAGCGAAAGGGAGGTAGGTATGTGGGATGAGTGAGGGGCCAGAGGGGAATTTCATTCCCT 120

Sbjct 1680174 AAGCGAAAGGGAGGTAGGTATGTGGGATGAGTGAGGGGCCAGAGGGGAATTTCATTCCCT 1680233





We've discovered that the animal is a dog! But what type of dog?

Article | Open Access | Published: 02 April 2019

Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology

[Jocelyn Plassais](#), [Jaemin Kim](#), [Brian W. Davis](#), [Danielle M. Karyadi](#), [Andrew N. Hogan](#), [Alex C. Harris](#), [Brennan Decker](#), [Heidi G. Parker](#) & [Elaine A. Ostrander](#) 

Nature Communications **10**, Article number: 1489 (2019) | [Cite this article](#)

26k Accesses | **94** Citations | **139** Altmetric | [Metrics](#)

Can you Identify the Dog's breed from its DNA?

set of genes an organism carries

TAAAAAGGATCT
TAAAAAGGATCT
TAAAAAGGATCT
TAAAAAGGATCT
TAAAAAGGATCT

Genotype

vs

Phenotype



physical characteristics of an organism



Determining which genotype causes each phenotype

	Long legs	ESR1
Beagle		

Beagle

ESR1

-----*

Beagle

ACGATCCGGGT A

GreatDane

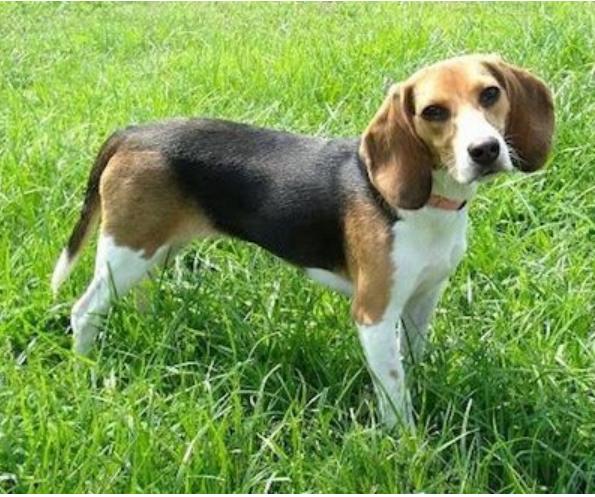
ACGATtCGGGTA

IrishTerrier

ACGATCCGGGT A

Finnish

ACGATCCGGGT A



Determining which genotype causes each phenotype

	Long legs	ESR1
Beagle	No	

ESR1

-----*

Beagle

ACGAT**C**GGGT

GreatDane

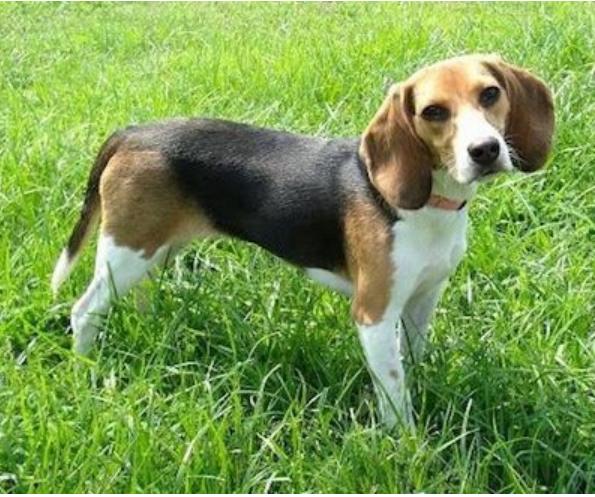
ACGAT**t**GGGT

IrishTerrier

ACGAT**C**GGGT

Finnish

ACGAT**C**GGGT



Determining which genotype causes each phenotype

	Long legs	ESR1
Beagle	No	C

ESR1

-----*

Beagle

ACGAT**C**GGGT

GreatDane

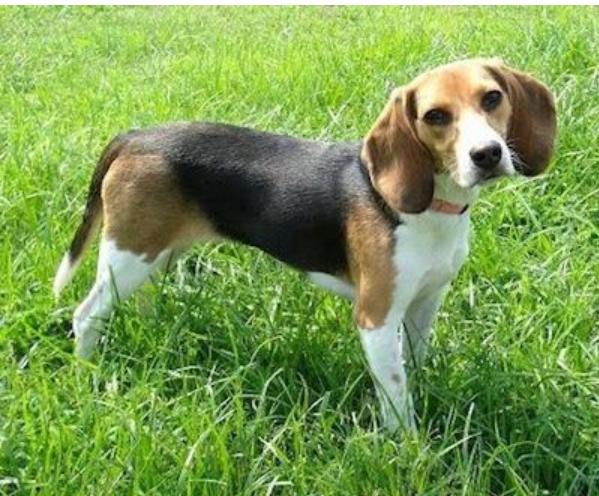
ACGAT**t**GGGT

IrishTerrier

ACGAT**C**GGGT

Finnish

ACGAT**C**GGGT



	Long legs	ESR1	Drop Ears	LncRNA
Beagle	No	C		

ESR1

-----*-----

Beagle ACGATCCGGTA

LncRNA

-----*-----

AGACGtTACAGC

GreatDane

ACGATTGGGT

AGACGtTACAGC

IrishTerrier

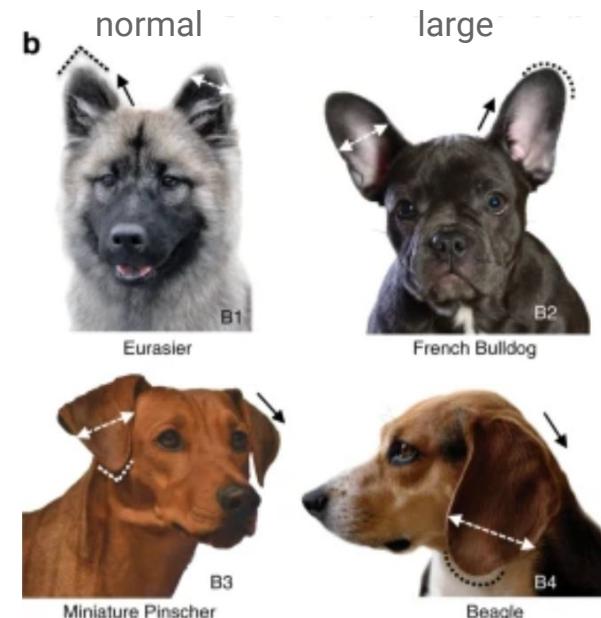
ACGATCCGGTA

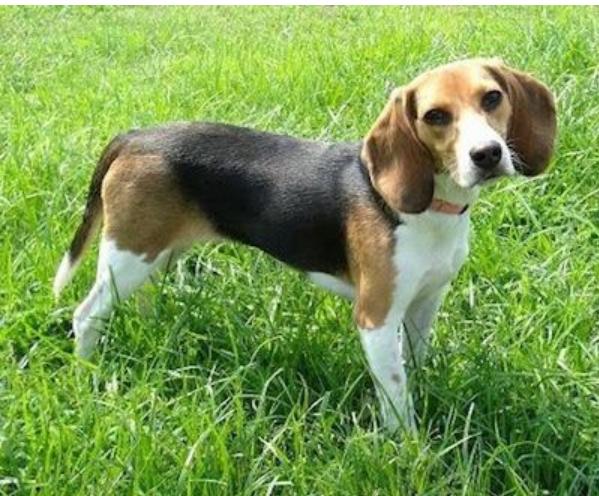
AGACGCTACAGC

Finnish

ACGATCCGGTA

AGACGCTACAGC





	Long legs	ESR1	Drop Ears	LncRNA
Beagle	No	C		

ESR1

-----*-----

Beagle ACGATCCGGTA

LncRNA

-----*-----

AGACGtTACAGC

GreatDane

ACGATTGGGT

AGACGtTACAGC

IrishTerrier

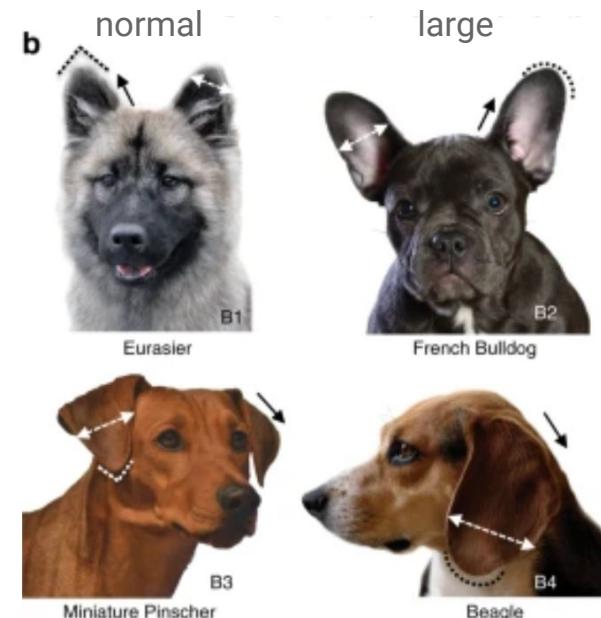
ACGATCCGGTA

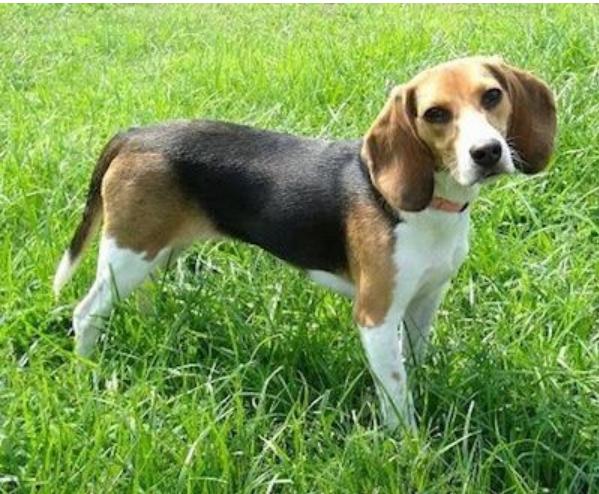
AGACGCTACAGC

Finnish

ACGATCCGGTA

AGACGCTACAGC





	Long legs	ESR1	Drop Ears	LncRNA
Beagle	No	C	Yes	

Beagle

ESR1

-----*

LncRNA

-- prick

Beagle

ACGATCCGGGT A

AGACGt TACAGC

GreatDane

ACGATT CGGGTA

AGACGt TACAGC

IrishTerrier

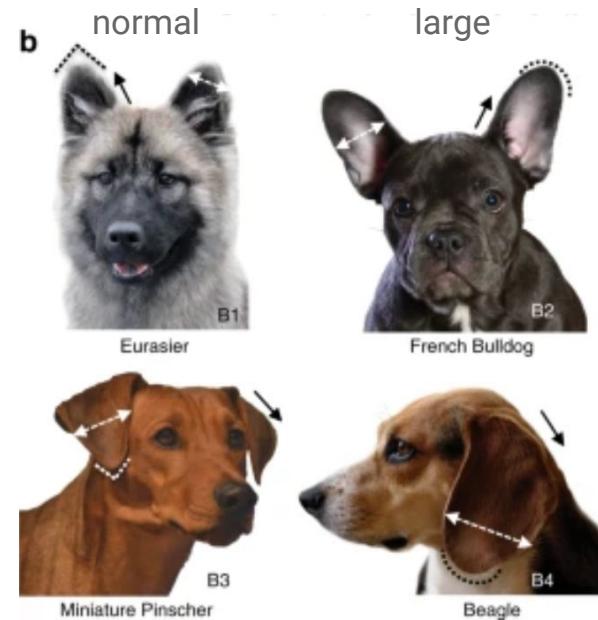
ACGATCCGGGT A

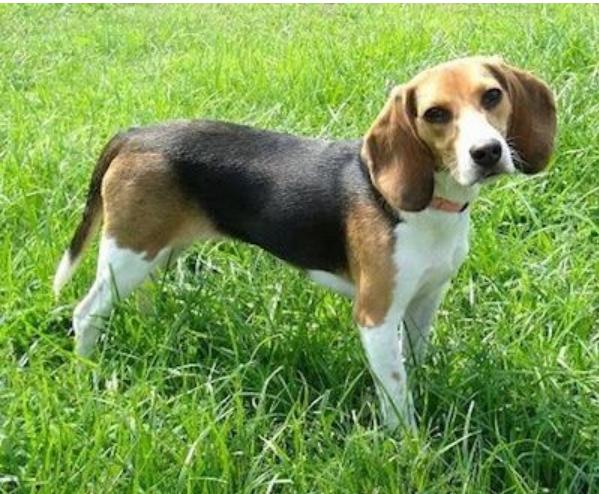
AGACGCT TACAGC

Finnish

ACGATCCGGGT A

AGACGCT TACAGC





	Long legs	ESR1	Drop Ears	LncRNA
Beagle	No	C	Yes	t

Beagle

ESR1

-----*-----

LncRNA

-- prick

Beagle

ACGATCCGGGT A

AGACGt TACAGC

GreatDane

ACGATT CGGGTA

AGACGt TACAGC

IrishTerrier

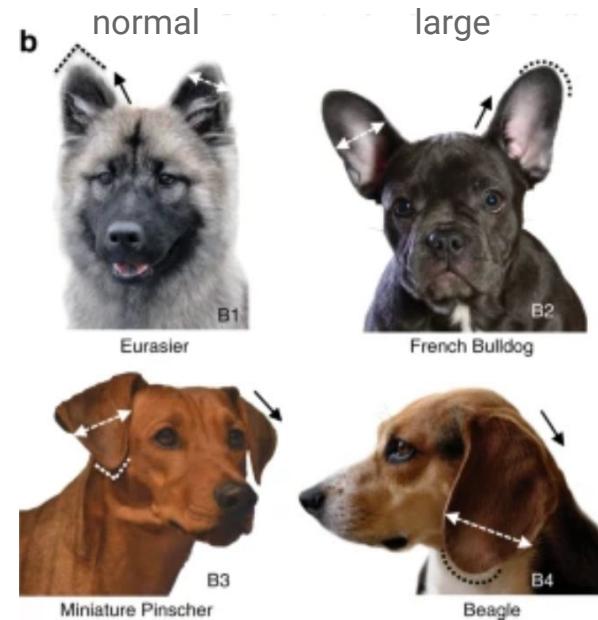
ACGATCCGGGT A

AGACGCTACAGC

Finnish

ACGATCCGGGT A

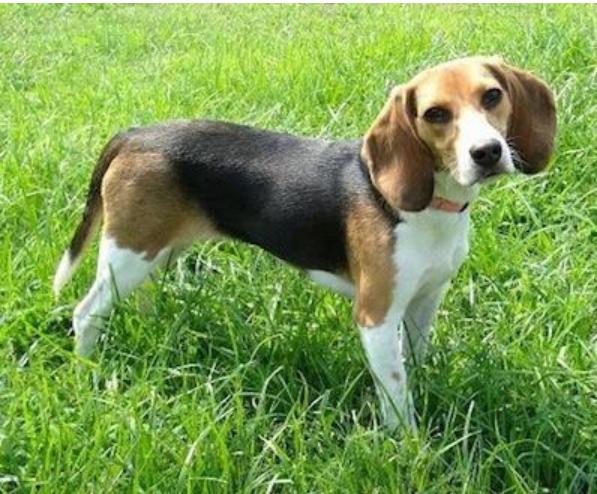
AGACGCTACAGC





	Long legs	ESR1	Drop Ears	LncRNA	Large ears	CFA12
Beagle	No	C	Yes	t	Yes	

	ESR1	LncRNA	CFA12	
	-----*	-----*	-----*	-----
Beagle	ACGATCCGGTA	AGACGTACAGC	CCAATCTAGAC G tATG	
GreatDane	ACGATT C GGTA	AGACGTACAGC	CCAATCTAGAC CC CATG	
IrishTerrier	ACGATCCGGTA	AGACGCTACAGC	CCAATCTAGAC CC CATG	
Finnish	ACGATCCGGTA	AGACGCTACAGC	CCAATCTAGAC CC CATG	



	Long legs	ESR1	Drop Ears	LncRNA	Large ears	CFA12
Beagle	No	C	Yes	t	Yes	t

	ESR1	LncRNA	CFA12	
	-----*	-----*	-----*	-----
Beagle	ACGATCCGGTA	AGACGTACAGC	CCAATCTAGAC GtATG	
GreatDane	ACGATT CG GT	AGACGTACAGC	CCAATCTAGAC CCATG	
IrishTerrier	ACGATCCGGTA	AGACGCTACAGC	CCAATCTAGAC CCATG	
Finnish	ACGATCCGGTA	AGACGCTACAGC	CCAATCTAGAC CCATG	



	Long legs	ESR1	Drop Ears	LncRNA	Large ears	CFA12	Mustache + eyebrows	RSPO2
Beagle	No	C	Yes	t	Yes	t		

	ESR1	LncRNA	CFA12	RSP02
	-----*-----	-----*-----	-----*-----	-----***-----
Beagle	ACGATCCGGTA	AGACGTACAGC	CCAATCTAGACGTATG	CATGCAT---GACTAC
GreatDane	ACGATTGGGTA	AGACGTACAGC	CCAATCTAGACGCATG	CATGCAT---GACTAC
IrishTerrier	ACGATCCGGTA	AGACGCTACAGC	CCAATCTAGACGCATG	CATGCATGGG GACTAC
Finnish	ACGATCCGGTA	AGACGCTACAGC	CCAATCTAGACGCATG	CATGCAT---GACTAC



	Long legs	ESR1	Drop Ears	LncRNA	Large ears	CFA12	Mustache + eyebrows	RSPO2
Beagle	No	C	Yes	t	Yes	t	No	-

	ESR1	LncRNA	CFA12	RSP02
	-----*-----	-----*-----	-----*-----	-----***-----
Beagle	ACGATCCGGTA	AGACGTACAGC	CCAATCTAGACGTATG	CATGCAT---GACTAC
GreatDane	ACGATTGGGTA	AGACGTACAGC	CCAATCTAGACGCATG	CATGCAT---GACTAC
IrishTerrier	ACGATCCGGTA	AGACGCTACAGC	CCAATCTAGACGCATG	CATGCATGGG GACTAC
Finnish	ACGATCCGGTA	AGACGCTACAGC	CCAATCTAGACGCATG	CATGCAT---GACTAC



Your turn!
Determine which genotype
causes each phenotype
Fill in table 1

Beagle



Great Dane

Finnish Lapphund



Irish Terrier



What breed is the mystery dog?

Fill in table 2

Determine what traits the dog has
Which of these dogs is it?



Charlie @SDHS - German Shepherd



Teddy -
Yorkshire
Terrier



Atlas - Bernese Mountain Dog

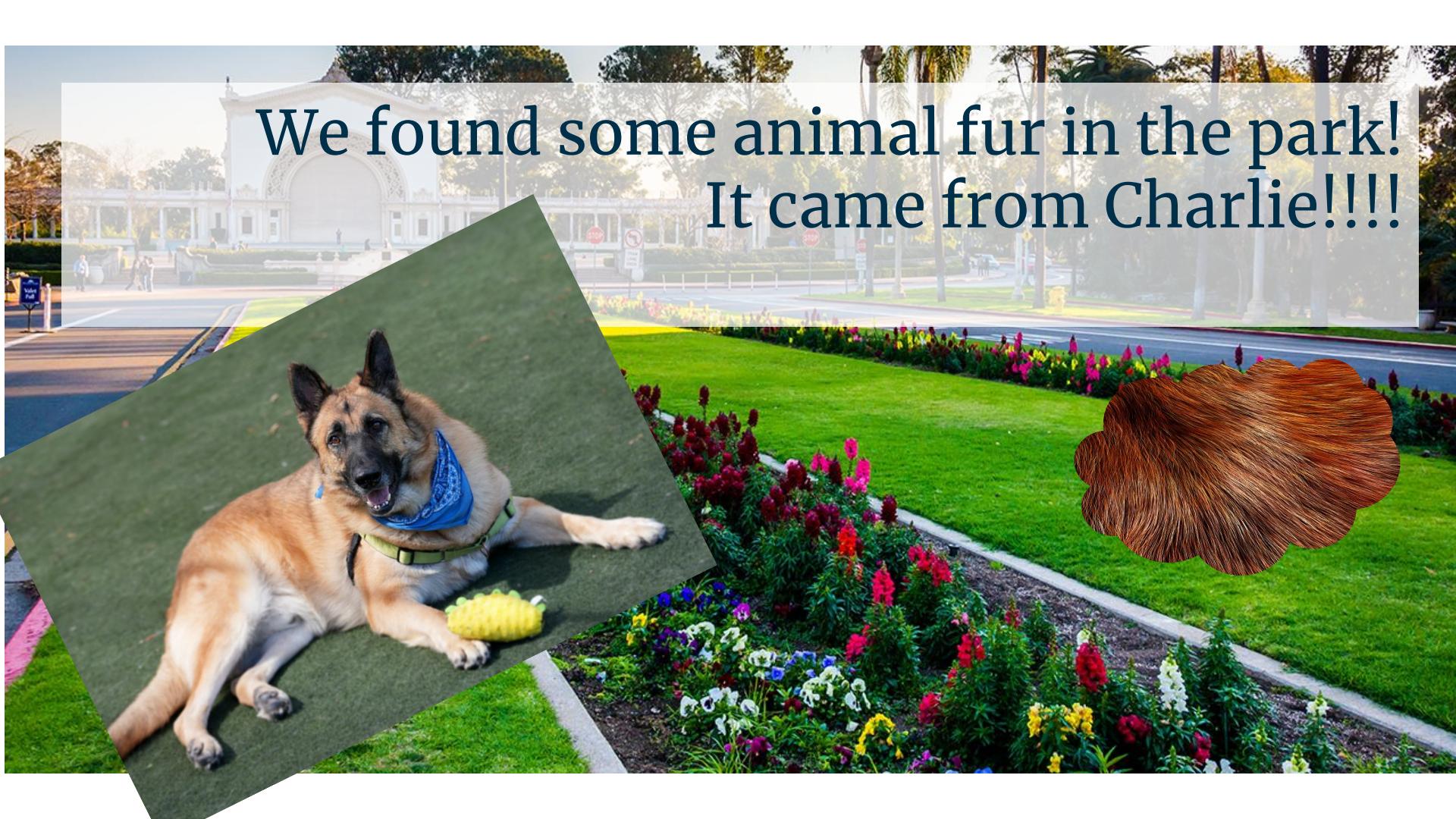
????

ACGATCCGGGTAA

AGACGCTACAGC

CCAATCTAGACGCATG

CATGCAT---GACTAC



We found some animal fur in the park!
It came from Charlie!!!!

