Dive into the microbial world and explore how DNA is unlocking this once hidden world!

Age: 11+

Genome challenge
Explore the big complexity of the smallest of creatures.
yourgenome.org/activities/genome-challenge

Extracting DNA from fruit
Extract DNA from a sample just like it’s done in the lab.
yourgenome.org/activities/extracting-dna-from-fruit

Function finders BLAST
Run some code cracking software and unlock nature’s potential from your computer!
yourgenome.org/activities/function-finders-blast

Construct a bug
Follow the DNA clues to discover if a bacterial sample is good or bad!
yourgenome.org/activities/construct-a-bug
A genome is an organism’s complete set of genetic instructions. All living things have genomes - each genome contains all of the information needed to build that organism and allow it to grow and develop. This includes the microbial world! There are many types of microbe including:

- **Bacteria**: good bugs and bad germs
- **Fungi**: microscopic moulds through to mushrooms
- **Protozoa**: single celled microscopic animals
- **Viruses**: parasitic pieces of DNA or RNA protected by protein shells

The Wellcome Genome Campus (home to the Wellcome Sanger Institute and EMBL-EBI) works in the field of ‘genomics’ (the study of genomes). This field has had a huge impact on the field of microbiology where often differences between species are incredibly hard to see without looking directly at the organism’s genetic code. Through this pack you will uncover how genomics can be used to explore the microbial world and how it is impacting on the field of biomedicine like never before.

**CHALLENGE 1: GENOME CHALLENGE**

Although every living thing has DNA, not all of it is the same. It is these differences in the genetic code that makes every species unique. Some of the most fundamental differences between groups of species is the size of their genomes and how many genes (individual DNA instructions) are contained within that genetic code.

In this first activity you are going to explore the genomics of four types of microbe and sort them by the size of their genomes and then the number of genes they have. Once you’ve had a guess check the answers and read about the scientific theories behind why they have the genomic features they do.
Malaria parasite  
(protozoa)

Coronavirus  
(viruses)

E. coli  
(bacteria)

Penicillium  
(fungi)
Put the organisms in order of genome size:

1. bigger genome
   (more DNA letters in the genetic code)

2. 

3. 

4. smaller genome
   (fewer DNA letters in the genetic code)
## GENOME CHALLENGE

### Worksheet

Put the organisms in order of number of genes:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

**more genes**
(more instructions in the genetic code)

**less genes**
(fewer instructions in the genetic code)
Put the organisms in order of genome size:

1. **Penicillium** (fungi) - 33 million base pairs
2. **Malaria parasite** (protozoa) - 23 million base pairs
3. **E. coli** (bacteria) - 5 million base pairs
4. **Coronavirus** (viruses) - 30,000 base pairs

**bigger genome** (more DNA letters in the genetic code)

**smaller genome** (fewer DNA letters in the genetic code)
Put the organisms in order of number of genes:

1. **Penicillium** (fungi) - 11,000 genes
2. **Malaria parasite** (protozoa) - 5,000 genes
3. **E. coli** (bacteria) - 4,000 genes
4. **Coronavirus** (viruses) - 12 genes

**more genes** (more instructions in the genetic code)

**less genes** (fewer instructions in the genetic code)
THE SCIENCE BEHIND THE ANSWERS

Fungi and protozoa tend to have larger genomes and more genes than bacteria and viruses. This is largely due to their cells being more complex. Both fungi and protozoa are eukaryotic, having similar cell structures to plants and animals. Bacteria are prokaryotes with less defined cellular compartments, for example they don’t have a nucleus, mitochondria or vacuole (found in plant cells). The size of the genome and number of genes shows a positive linear correlation with the larger the genome the more genes the species tends to have. Interestingly, this trend is not seen in eukaryotes.

Viruses are not strictly ‘living things’ when usual tests of life (such as their ability to reproduce independently of a host, having the ability to self-regulate their internal cell environment, etc.) are applied to them. They are essentially short pieces of DNA wrapped in a protein coating. Like bacteria they tend to show a positive linear correlation between genome size and number of genes although they typically have significantly smaller genomes and far fewer genes. This is largely due to their parasitic lifestyle – utilising the cellular processes of the host they infect to propagate, rather than having the genetic code for those processes themselves.

**Penicillium (Penicillium chrysogenum)**

This is a species of fungi that is often found in damp and water damage within houses. It can also be found on salted foods and is rarely linked to human disease. As with other species of the Penicillium genus, it is used in industry to make the antibiotic penicillin and numerous other antimicrobial compounds for use in medicine. It has the ability to make these products as it uses them to outcompete bacterial species in its natural environment.

Stats rounded to 2 significant figures and correct as of 28/05/2020: fungi.ensembl.org/Penicillium_chrysogenum_gca_000710275/Info/Annotation/
Malaria parasite (*Plasmodium falciparum*)

Protozoa are single celled eukaryotic organisms. They were historically likened to microscopic animals due to having similar characteristics such as the ability to move and feed on other organisms. The genus of protozoa *Plasmodium* causes the disease malaria with *P. falciparum* being the deadliest of the species in this genus to humans. It is parasitic, damaging its human hosts’ blood and liver cells to survive and propagate. Humans are only one step in the complex life cycle of the parasite with a mosquito host (*Anopheles gambiae*) playing a vital role in transmission between human hosts.

Stats rounded to 2 significant figures and correct as of 28/05/2020: protists.ensembl.org/Plasmodium_falciparum/Info/Annotation/#assembly

**E. coli** (*Escherichia coli*)

Even within a bacterial species there can be multiple different strains of the bacteria. This is largely due to shorter generation times which leads to greater variation arising far quicker than in eukaryotic organisms. *E. coli* strains have been used for many years as a model organism for lab research due to the ease of altering their genetics and growing in large quantities, making them ideal for biomedical and industrial applications. Most strains are harmless to humans however some can cause infections such as serious food poisoning.

Stats rounded to 2 significant figures and correct as of 28/05/2020: bacteria.ensembl.org/Escherichia_coli_str_k_12_substr_mg1655/Info/Index

**Coronavirus (SARS-CoV-2)**

Viruses are small packages of DNA (or RNA) held within a protein coat. They infect host organisms and use the host’s cellular mechanisms to reproduce and spread. There are many species of coronavirus that cause respiratory disease in humans and animals. They are grouped together because of the shared crown-like spike proteins found on their surfaces. These proteins help the virus to adhere to and infect a host’s cells and as such are a focus of research for treatments and vaccine development. SARS-CoV-2 is the coronavirus responsible for the COVID19 disease.

Stats rounded to 2 significant figures and correct as of 28/05/2020: covid-19.ensembl.org/Sars_cov_2/Info/Annotation
No matter the organism, knowing how big the genome is and how many genes it has is only the first step! By reading the DNA code, we can work out what proteins the genes code for and what they do within the microbe.

**CHALLENGE 2: EXTRACTING DNA**

Let’s now look at what DNA – the molecule in which all genetic instructions is written – really looks like. You can find DNA in every living organism and to research an organism’s genomics you first have to get the DNA out of the cells!

Follow the instructions on page 11 to see how scientists extract DNA from a sample. For this activity we will be using strawberries as they are normally easy to find in shops but any soft fruit works great for the experiment! The process is very similar to that used in labs when working with microbial samples.
EXTRACTING DNA FROM FRUIT

Instructions

1. Mix the washing-up liquid, the water and the salt in one cup.

2. Stir gently to mix it all together.

3. Put the fruit in a bag and add the mixture.

4. Squeeze the air out and seal the bag.

5. Squash to break up cells and release the DNA.

6. Pour the fruit mixture through the sieve and into the jug.

7. Half fill a small cup with the liquid from the jug.

8. Slowly pour very cold alcohol down the side of the small cup.

9. You should see white clumps forming in the clear layer... that's DNA!
How much DNA did you get? It can be tough to hook the DNA out on a toothpick – some scientists spend years getting good at it! Because DNA is such a compact molecule, there is actually enough DNA in your cup that, if stretched out into a single strand, could go all the way from the UK to Hawaii!

By extracting the DNA from a microbe and “reading” it using DNA sequencing machines scientists can identify the different proteins produced by microbes and can help us work out what they do. Proteins are the building blocks of life and carry out many different functions - in the case of a microbe, some of the proteins they produce are toxins that make us ill, others produce enzymes that can break down plastic!

Studying these gene sequences and the proteins they produce can help us understand the causes of disease and identify potential new treatments.

**CHALLENGE 3: FUNCTION FINDERS BLAST**

Once you have the DNA code of a microbe you need to work out what those genetic instructions are producing in the microbe. No matter the organism, the process of turning DNA into proteins is the same fundamental process.

Your next task is to decode some DNA sequences to find out what protein they produce. Use the codon wheel on page 13 to translate the DNA codes on your worksheet into a protein sequence. Once you have your protein codes, follow the instructions to carry out a BLAST search using the UniProt database where you will find out what your protein is and what it does.

When you have completed the challenge, think about these questions, and if you can discuss them with your teacher, friends or family:

- What was the most interesting protein you found?
- Do you think you could use this research for understanding how specific microbes work?
Use the codon wheel to translate DNA codons into amino acids:

To decode a codon find the first letter of your sequence in the inner circle and work outwards to see the corresponding amino acid. For example: CAT codes for H (Histidine).

Please note that this wheel uses the sense DNA codons (5’ to 3’).
Translate the DNA sequences to find the matching protein using Uniprotein BLAST search:

| DNA sequence 1 | atg  aag  tca  gct  att  tta  acc  ggt  ttg  ctt  ttc  gtc |
| Translated sequence | __  __  __  __  __  __  __  __  __  __  __  __ |
| Protein name | __________________________ |
| Organism | __________________________ |
| Protein function | __________________________ |

| DNA sequence 2 | atg  agt  aaa  gga  gaa  gaa  ctt  ttc  act  gga  gtc  gtt |
| Translated sequence | __  __  __  __  __  __  __  __  __  __  __  __ |
| Protein name | __________________________ |
| Organism | __________________________ |
| Protein function | __________________________ |

| DNA sequence 3 | gaa  aac  atg  gag  aac  gat  gaa  aat  att  gtg  tat  ggt |
| Translated sequence | __  __  __  __  __  __  __  __  __  __  __  __ |
| Protein name | __________________________ |
| Organism | __________________________ |
| Protein function | __________________________ |

| DNA sequence 4 | ggt  tgg  gct  ttg  cgg  atc  atg  ttt  cta  cat  ctg  tac |
| Translated sequence | __  __  __  __  __  __  __  __  __  __  __  __ |
| Protein name | __________________________ |
| Organism | __________________________ |
| Protein function | __________________________ |
Translate the DNA sequences to find the matching protein using Uniprotein BLAST search:

<table>
<thead>
<tr>
<th>DNA sequence</th>
<th>Translated sequence</th>
<th>Protein name</th>
<th>Organism</th>
<th>Protein function</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>cct ggg gag aac cta tgc tat aga aag atg tgg tgc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ccc aga gag atc cag acc gcc gtg aga ctg tta ctc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>gag aag aga aag ctg ttt atc cgt tcc atg ggt gaa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>atg gag ttt act ttg agg caa gag gct tta gtt aat</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FUNCTION FINDERS BLAST
Worksheet instructions

HOW TO COMPLETE THE WORKSHEETS

1. Use the codon wheel to **translate the DNA sequences** on the worksheet to amino acids.

2. **Type the amino acid sequence** in to the Uniprot Blast search [www.uniprot.org/blast/](http://www.uniprot.org/blast/).

   ![BLAST search interface]

   Press **Run BLAST** to get results (it may take a few seconds for results to appear).

3. When the search results appear, **filter the results** to only show reviewed entries (gold file icon with a star). Each result is known as a “hit”.

   ![Filtering results in BLAST]

   ![Alignment view in BLAST]
FUNCTION FINDERS BLAST
Worksheet instructions

4. After filtering the hits should look like this. **Scroll down** to the Overview section.

Look at the info column. This will give you an idea of how reliable your hits are. The Expect value (E-value) indicates the number of random hits you would expect by chance for the given query sequence and the size of the sequence database against which the BLAST is performed.

For example, an E-value of 1.0 means that you would expect on average to get one match in the database for the submitted query simply by chance. The lower the E-value, or the closer it is to zero, the more “significant” the match is. In general, the E-values should be in the range of 0.01 to 0.1 to be statistically significant.

The identity % describes how similar your sequence is to the hit, i.e. whether the amino acids are in the same position when aligned. 100% means the sequences match exactly.
5. **Click on the hit** that you think best matches your sequence. Find out the name of the protein which species the sequence is from (common and species name) and what the protein does. Is the protein found in other species?

You can reduce the amount of information on the screen by unticking the blue display categories on the left hand side of the screen.

If you cannot find all the information you need, try using a Google search or Wikipedia to find out more.

6. To start a new BLAST search click **Edit and resubmit**, and enter your next set of amino acids. Repeat steps 3 to 5.
Decoding DNA, in the way you have done, is a powerful tool but only really works if you have already researched what the specific genes and proteins do.

Sometimes changes can occur within a genome as a result of naturally occurring mutations or a response to an environmental stress (like an antibiotic). These changes are small, just one or two DNA letters, so when studying microorganisms, you can’t find an exact DNA match - find the closest match. This can help you to work out what a particular gene does.

This approach is really useful when working with “new” microbes that have never been seen before and can help scientists to work out if that microorganism could make someone seriously ill, if infected with it.

**CHALLENGE 4: CONSTRUCT A BUG**

In this task you will be taking on the role of a clinical scientist sequencing an unknown bacterial sample from a patient. Start by selecting six different DNA reads at random from the sequence reads sheet. These will be the DNA codes you’ve managed to sequence from your sample.

Now try and match them to the DNA codes on the genome reference sheet. If they match then it means your sample had that same DNA code present and more than likely that same gene! Each of the genes are named and represented by arrows above the DNA code. As you find where your codes match, tick the genes off on your gene profile card.

When you’ve finished, think about what a bacteria with these genes might be like and answer the questions on the worksheet. You can compare this with known bacteria profiles to find out what sort of bacteria you have and what it does.
CONSTRUCT A BUG
Sequence reads

ATAAAGGTCCGGAATTTGCCCAGCTT
AGGAGCGCTTCCC
TGATCTCCGCCATCAACTCGTACGCTG
AGTTACCGCGCTTACAGACGCT
ACAGGACAAAGATTATACAGACATTTC
GGTTAACGACAACATACGTTGC
CGTAAAGGAGCG
CATATG
TTAAGCGACATT
GCGCTGAG
ATACGGAAAATATCAGATACAG
GGATCTTAGCTAAAGGTCC
ATCTCA
GGTGATCTCAGCATCAACTCGT
GTCAGGCTATCTGGATGACGTC
GATGATGATGTTTATTAAA
TATTACTGACATGGACGCAT
GTAAAGGTCCCG
AATTTGCCGC
AGAAAATG
GCAAAAG
GTTCAGCTGCTCTCATGATG
ATGACAAAC
TGTATGATGATGATGAT
TGATACATCG
CTAAAGGTC
TAAAGGCC
ACGATCCATCTCCAG
GCCAAACATAT
ATCATTTT
CONSTRUCT A BUG
Reference genomes

**Bacteria 1 – Klebsiella pneumoniae**

- **KPC-2**
  - AAGTTACCGCGCTGAGGAGCGCTTCCCACTGTGAGCTCA
  - AGTGATCGTGATAGCGCAGGGAATTAACGAGACATTATATTCC

- **MprA**
  - GCCAAACATATGGCATGGGTTAACGACACCAATACCGTTGCGGAAGACTCCGTTGCCACTGGCAATGCTCGCGTAAAGGAGCG

- **LuxS**
  - GCAGAGGAAAAATGGTGACGATCGCCATCTCCACGGTGATCTCCGCGCATCAACTCGCTGACGCTCGAGATCTAGCTAAAGGTCC

- **OqxB**
  - ATACGGAAAATATCAGAGTACAGGACCAAAAGATTTTACAGACATTTCTCCATGCAAAGGATTATTACTATGACATGGACCATCTTC

- **AtpA**
  - GCCAACACATATGGCATGGGTTAACGACACCAATACCGTTGCGGAAGACTCCGTTGCCACTGGCAATGCTCGCGTAAAGGAGCG

- **CcpA**
  - ATGACAAAAACTCACCGCTAGTTTTGCTGCTTTCATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

**Bacteria 2 – Shewanella oneidensis**

- **LuxS**
  - GCCAAACATATGGCATGGGTTAACGACACCAATACCGTTGCGGAAGACTCCGTTGCCACTGGCAATGCTCGCGTAAAGGAGCG

- **BpfA**
  - ATGACAAAAACTCACCGCTAGTTTTGCTGCTTTCATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

- **CcpA**
  - ATGACAAAAACTCACCGCTAGTTTTGCTGCTTTCATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

- **FlhA**
  - TTTTCTTATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

- **MtrC**
  - TTGCGCTTTCATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

- **MexF**
  - TTTTAATTAAAAAGACCCGA

**Bacteria 3 – Salmonella typhi**

- **LuxS**
  - GCCAAACATATGGCATGGGTTAACGACACCAATACCGTTGCGGAAGACTCCGTTGCCACTGGCAATGCTCGCGTAAAGGAGCG

- **AtpA**
  - ATGACAAAAACTCACCGCTAGTTTTGCTGCTTTCATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

- **FlhA**
  - TTTTCTTATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

- **TviA**
  - TTTTCTTATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

- **Prgl**
  - TTTTCTTATGATGAAATTCGTGGAATGTTAGTATGATGATGATGTTTTATTAAAAGACCGA

- **SptP**
  - ATACGGAAAAATATCAGAGTACAGGACCAAAAGATTTTACAGACATTTCTCCATGCA

- **IpfC**
  - AAGATTATTACTATGACATGGACGCATCTTTC
## Environmental protection genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TviA</td>
<td>Involved in capsule production protecting bacteria from environmental factors</td>
</tr>
<tr>
<td>MprA</td>
<td>Involved in capsule production protecting bacteria from environmental factors</td>
</tr>
<tr>
<td>SptP</td>
<td>A protein released into hosts that suppresses the immune system</td>
</tr>
</tbody>
</table>

## Toxin defence genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPC-2</td>
<td>An enzyme that breaks down penicillin based antibiotics</td>
</tr>
<tr>
<td>CcpA</td>
<td>Removes toxins produced from radiation</td>
</tr>
<tr>
<td>OqxB</td>
<td>Removes chemical toxins from the bacteria</td>
</tr>
<tr>
<td>MexF</td>
<td>Pumps chemicals like antibiotics out of the bacteria cells</td>
</tr>
</tbody>
</table>

## Lifestyle genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MtrC</td>
<td>A molecular wire that discharges electrons (essentially making electricity)</td>
</tr>
<tr>
<td>AtpA</td>
<td>Part of the system that releases energy for cells to live and function</td>
</tr>
<tr>
<td>PrgI</td>
<td>Forms a needle-shaped structure that injects effector proteins into host cells</td>
</tr>
</tbody>
</table>

## Community genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlhA</td>
<td>Makes flagella (a tail) that allows the bacteria to move in its environment</td>
</tr>
<tr>
<td>IpfC</td>
<td>Enables bacteria to colonise the gut</td>
</tr>
<tr>
<td>LuxS</td>
<td>Allows bacteria to sense how many of them there are in an area (quorum sensing)</td>
</tr>
<tr>
<td>BpfA</td>
<td>Helps bacteria stick together in biofilms</td>
</tr>
</tbody>
</table>
1. Look at the genes your bacteria has. What adaptations or features do they give it and how do these contribute to its survival strategy?

2. How might a bacteria become more adapted to a pathogenic lifestyle?

3. Which of the 3 bacteria (*Klebsiella, Shewanella & Salmonella*) do you think are most closely related?

4. Which genes do you think are the most essential / oldest? Hint: older and more essential genes are usually retained in many species as they are critical to their survival.
### Bacteria 1

**Klebsiella pneumoniae**

- A human pathogen (causes disease)
- Often resistant to antibiotics
- Can sense how many of the bacteria are present (called quorum sensing)
- Can use quorum sensing to alter behaviour during infections to better survive

### Bacteria 2

**Shewanella oneidensis**

- This bacteria do not cause human disease
- Adapted to live in soil and river sediments
- Often forms biofilms with other types of bacteria to survive
- Ability to generate electrical currents makes it useful for biotech and renewable energy

### Bacteria 3

**Salmonella typhi**

- A human pathogen (causes disease)
- A cause of food poisoning
- Injects proteins into host cells to damage them
- Uses tail-like structures (flagella) to move
- Has a number of systems that allows it to survive in and colonise the gut during infections
An amazing feature of bacteria is their ability to share and swap sections of DNA. Constructing a bug, like you just have, is not so far from reality as you might think. For example, this is how antimicrobial resistance (the ability of microbes to survive antibiotics designed to kill them) arises in populations. A small group of bacteria will have genes that give them resistance to an antibiotic and so thrive when competitors are killed by the drug. They can also pass those protective genes to other bacteria (often ones that they symbiotically live with) so that the other bacteria can also survive and thrive.

The microbial world is all around us. Wherever you look there are microbes living even if you cannot always see them! Genomics is providing us with more tools to explore that hidden world allowing us to make many discoveries affecting a range of scientific fields. From offering clues to new treatments for disease, making food, recycling plastic and helping us to understand how life on Earth began – microbes are some of the most amazing creatures on Earth!

WHAT DO YOU THINK?

Now that you have finished the pack have a go at answering these questions!

www.surveymonkey.co.uk/r/Genomics_and_Microbes