

Finding a Gene in DNA

▶	What is a String	2 min
▶	Understanding Strings	3 min
▶	Developing an Algorithm	5 min
▶	Positions in Strings	8 min
▶	Translating into Code	11 min
▶	Java Math	8 min
📋	Programming Exercise: Finding a Gene and Web Links	10 min
★	Practice Quiz: Finding a Gene in DNA	6 questions

Finding All Genes in DNA

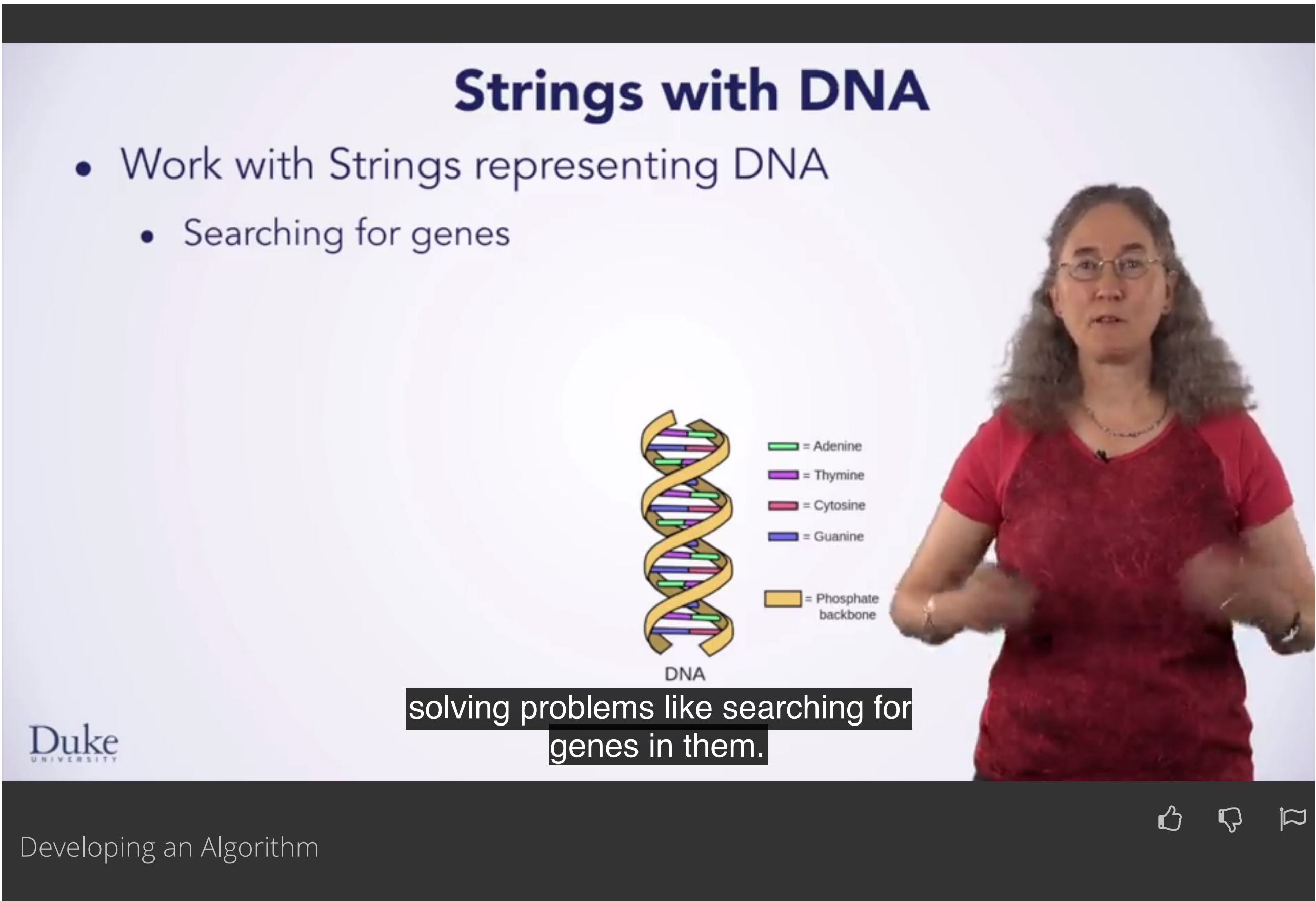
Debugging Code

Using the StorageResource Class

Review

Strings with DNA

- Work with Strings representing DNA
 - Searching for genes



Developing an Algorithm

Have a question? Discuss this lecture in the week forums. >

Interactive Transcript

Search Transcript

English ▼

0:03
Hello and welcome back. You are going to be working with strings that represent DNA, [solving problems like searching for genes in them](#). This problem is a great one to learn on because, even though we're going to start with a greatly simplified version of it, it's an important problem with real world applications.

0:23
Of course, the lessons you will learn about working with strings and programming in general will help you far beyond this problem domain. Whatever sorts of problems you want to solve, strings are likely to come up in them in one way or another. HTML, email, or really anything that is text that is represented as a string. You are also going to learn some other important lessons as you work on these problems, like how to do math in Java. Of course, math is also ubiquitous in programming since everything is a number. And for that most importantly, you will get more practice developing and implementing algorithms with this seven steps. Before we dive into DNA related problems we need to give you a bit of domain knowledge, some terms and concepts related to working with DNA.

1:14
Here is a string that could represent some piece of DNA. You will see that it's made up of four letters A, T, C and G.

1:23
Each of these represents one nucleotide Which are the basic building blocks of DNA.

1:29
Three nucleotides together make a codon, which each describe one amino acid.

1:36
The ATG codon shown here is special in that it indicates the start of a gene. Accordingly, it's called the start codon.

1:45
And the TAA codon is also special in that it indicates The end of a gene, so it's called the stop codon.

1:52
There are a couple other stop codons, but for now, we're only going to think about TAA.

1:59
Everything between and including these two codons makes up one gene.

2:04
The first problem you're going to work on is finding a gene in the string which represents DNA. That is you want to write a program which takes a string like this one and gives you all the text between it and including the start code on ATG and stop code on TAA. You are going to start with a greatly simplified version of this problem. Just finding those letters. And all the text between them.

2:32
You will not worry about the fact that real genes must be multiples of three in length because they're made up of codons. Or that there are some other stop codons. Or a few other complexities to start with. As you master more string and algorithm concepts you'll add features to your program. Making it more realistic with each step.

2:53
As always, the first thing you want to do is work on an instance of the problem yourself.

2:59
Let's take this DNA sequence, and find the first gene in it.

3:05
Let's find the start codon. Aha! There it is! Now let's start looking after it to find the stock code on. Then we find it right here.

3:17
That means that we have to take all this text representing the nucleotides in this region as our answer the gene that we found

3:29
Now that we have worked an example, we should write down what we just did.

3:37
First, I found the first occurrence of ATG.

3:42
Then I started looking after the ATG for TAA Last, I took all the letters between, and including them as my answer, ATGATTTTC etc., all the way to TAA.

4:01
Okay, now that we wrote down what we did for that specific problem, we want to generalize it.

4:08
Why did we look for ATG?

4:11
We always want to look for it. That's the start codon.

4:14
What about TAA after ATG?

4:18
We always want to do that, too. That is the stop codon.

4:23
and taking all the letters between and including them? We always want to do that too.

4:29
The only thing that isn't really general here is the specific string we wrote down as our answer, which was more of a descriptive note to ourselves than anything else. Now that we have a general, we'd like to, learn

4:57
java script first And how do we get all the letters in a particular range in the string?

5:06
You'll learn about these concepts then you'll be ready to turn the algorithm into code, thank you.

Downloads

Lecture Video	mp4
Subtitles (English)	WebVTT
Transcript (English)	txt

Would you like to [help us translate](#) the transcript and subtitles into additional languages?