

## Decrypting a Text (1964 letters) Encrypted by Substitution

Here is a 1964-character text encrypted by substitution (non-letters removed). We know that the plain text was English language text. We decrypt it by frequency analysis.

RIRPKQJHIRBTLCSWQKUXBVERIVEHUZCCIGIVQJWKGHJCSCUVLUGHVBSCSLKVXVQBPJRLHURBTKTUH  
EGRKJRLWHWVLKIGHWQJHCLUGHQLCUHWKURUHKRUUGHHWLWXRCNCNTHRECKKCVLUGHKURBWQKUBLI  
RPKQJHKJVZHDXBVEEPGVXHHWPBKHVILWRKCUHUQBLHWRUREERBDCLSUGHXCBKUKQIIHKKXQJVJJH  
IUCVLVXIVEHURBTBLWCLUHBKUHJJRBERUHBCRJRLWWBRZCLSIGHHBKBVELRKICHLUCKUKOVHMHJJC  
LSRRLWKHMHBRJWVYHLLIVJHRSQHKSRRUGHBHWRLNCVQKJTTKUHBWRURECJCURBTUHKUCLSLSHCIQ  
URGZRUIGCLSRKUGHCPBFVHFUGHKCYHVXRJRBSPJRLUPVUGQBUJHWCLUVUGHRUVEVKPGHBXRKUHBGRRL  
TERLERWHVFOHIUGRWVVLHFHVXBHRKIVEHUKRBHJRBSPFRJJJKVXWCUTCIHJXUVMBXBVEUGHIBHRUCV  
LVXUGHKVJRBKTKUHEFCJJCVLTHRBRKSVKICHLUCKUKHNPHIUUGRUGHKREPJHKZCJBBMHRIJIQHKRFV  
QUGVZUGHPJRLHUKZHBHXVBWEHWRWPVKCFJTUGHVBSCSLKVXJCXHEBMHJJCLSRLHBMVQWKCPVKUCUV  
LZRKLVUKVJHJTFHIRQKHXVUGHGCSKGPKHHWIQJECLRUCVLVXUGHECKCVLVBZGCIGGHCKPBVSBREEHER  
LRSHBXVBJVIDGHHWERBUCLKPRIHKTKUHEKUGHIVEPRLTUGRUWHKCSLHWRLWVLKUBQIUHWUGHPBVFHRL  
WCUKEVUGHBKGCPKCNUHHLEVLUKGKRSVGHZRUGHWCLGVBVBRKUGHSHLHKCKBHUQBLPBFVHIRBTLCSKV  
JRBZCLWPRBUCIJHKFQCJUFTUGHKREHUHREQLWHBGCKWCBIUCVRLWFRKHVLKCECRJBUHIGLJVSTKJ  
REEHWCLUVUGHGSBVQLWRUEPGHMHBTFVWTCKIVLXCWHLJULZGRUZHGRMHWKCSLHWRLWFQJCJURLWUHKUHW  
GHKRCWFQHMHBTFVWTTRJZRTKZVBBCHKRFPVQZGRUZHGRMHVLUVUGVQSGUVXKURBWQKUZRKJRLIGHWCLX  
HFBQRBTCUHLIVQLUHBHWUGHZCJWIVEHUUZVTHRBKRSVRLWGVPHQJUTSRUGHBHWQKUXBVECUKURCQJK  
CLSRPRWWJHKGRPHWJCDHRUHLCLCKBRIAQHURHBVSHJRIJVQBJHKKSRKVQKOHJJTUGRUCKUGHJCSGUHK  
UERUHBCRJVLHBRUGRLWCKDLVZLUVKICHLUCKUKRKKVYHLKEVDHZRKQKHWCUGHRUUHEPUUVUBRPUGHE  
RUHBCRJWBKRUUSHLSHXBVECEPHCRJIVJJHSJVLWVLJHRCLSHNPBHVULHNUBRUBBBHKUBCRJWQKUZ  
CJJFHVLVXUGHXCBKUFBCUCKGKICHLUCKUKVUBHIDCMHWQKUKREPJHKUVRLRJTKHUGCKCUGHXCBCUUC  
EHKCLIHUGHRPVJJVECKKCVLKUGRKREPJHKVBVIDGRMFHHLBHUQBLHWXBVEKPRIHUVHRBUGGHKRCWU  
GCKUGVQKRLWUGVXRSBREEHVXQKUXBVEZCJWZCJJPBVRFTJUHJQKEVBHRFVQOUGHVBERUCLVXUGH  
KVJRBKTKUHEUGRUGHPRKUTHRBKVXUHJKIVPHVFKHBMRUCLVKVXUGHKHFVFOHIUKCUCKRSBHRUUCEHUV  
FHCLUVWQKUERLTFHJCHMHUGRUIVEHUKWHJCJMHBHWEVKUVXUGHZRUHBLHRBUGRLWPVKCFJUTGHVBSRL  
CIIVEPVQLWKUGRUGZHBHMCURJUVUGHMVJQUCVLVXJCXH

**1) It is very likely that  $e \rightarrow H$ ,  $h \rightarrow G$ ,  $t \rightarrow U$ .**

Look at the frequencies of common plaintext and ciphertext letters and digrams, all expressed in occurrences per 10000 characters of text. Plaintext frequencies are estimated using about 3.2 million characters of what is (hopefully) fairly typical English language text. (Recall lower case indicates plaintext, upper case indicates ciphertext).

Letter	Frequency (per 10,000)	Letter	Frequency (per 10,000)	Digram	Frequency (per 10,000)	Digram	Frequency (per 10,000)
e	1237	H	1253	th	330	UG	290
t	921	U	1013	he	302	GH	250
a	821	R	845	an	181	KU	219
o	767	K	804	in	179	HK	163
n	705	V	743	er	169	RU	158
i	676	C	642	nd	146	UH	153
h	645	B	616	re	133	HB	148
s	617	L	580	ed	126	RL	143
r	550	J	484	es	115	HW	138
d	479	G	479	ou	115	VL	133
l	393	W	397	to	115	CL	133
u	291	E	316	ha	114	RB	133
		I	280	en	111	UC	107
		Q	260	ea	110	LW	107
		X	229	st	109	HU	102
		S	204	nt	106	HR	102
				on	106	VX	102
				at	104	BH	102
						CK	102

Merely from the single-letter frequencies, it appears very likely that  $e \rightarrow H$ . Otherwise the plain text character  $\lambda$  for which  $\lambda \rightarrow H$  would have  $freq(\lambda) \leq 921$ , versus  $freq(H) = 1253$ . So we assume  $e \rightarrow H$ . In the unlikely event this is wrong, we will discover our error, and backtrack.

The highest-frequency plaintext digram **th** ( $freq = 330$ ) is very likely mapped to one of the highest-frequency ciphertext digrams: **UG** ( $freq = 290$ ), **GH** ( $freq = 250$ ), or **KU** ( $freq = 219$ ). But **GH** is impossible because  $e \rightarrow H$ , and **KU** is unlikely because  $freq(U) = 1013$  is way higher than  $freq(h) = 645$ . So we assume **th**  $\rightarrow$  **UG**. In particular **t**  $\rightarrow$  **U** and **h**  $\rightarrow$  **G**.

Comparing frequencies of characters we have decrypted so far, we have

plaintext	e	t	h	ee	tt	hh	et	te	eh	he	th	ht
frequency	1237	921	645	48	56	6	83	75	33	302	330	32

ciphertext	H	U	G	HH	UU	GG	HU	UH	HG	GH	UG	GU
frequency	1253	1013	479	41	56	10	102	153	15	250	290	20

Agreement is quite good (better than we might expect) except in two entries:

- i) The frequency **G** is rather low, compared to **h**.
- ii) The frequency of **UH** is surprisingly high, given that of **te**.

However, even in these cases the deviations aren't as large as those we used above to reject alternatives (e.g., to reject **t** → **H** or **h** → **U** or **th** → **RU**).

**2) It is very likely that  $n \rightarrow L$ , and  $\{a, o, i\} \rightarrow \{R, V, C\}$  in some order.**

In our plaintext data, the third and fourth most common digrams are **an** ( $freq=181$ ) and **in** ( $freq=179$ ). In addition, **en** ( $freq=111$ ) and **on** ( $freq=106$ ) are very common. On the other hand, **tn** ( $freq = 7$ ) and **hn** ( $freq = 1$ ) are quite rare.

We don't yet know how to encrypt **a**, **i**, **o**, or **n**, but they are among the six most frequent plaintext letters, so we expect them to encrypt to high-frequency ciphertext letters. It is probably safe to assume that **a** and **o** encrypt to a letter **L** or above, in the list of ciphertext letters sorted by frequency, and **n** and **i** to a letter **G** or above. Excluding **H**, **U**, and **G**, this means each of **a**, **i**, **o**, **n** must encrypt to ones of **R**, **K**, **V**, **C**, **B**, **L**, **J**.

Consider the following table of digram frequencies below. The first letter of the digram is given at left, the second letter at top.

Alternatives for the letter that <b>n</b> encrypts to								
What should occur in the column for the letter that <b>n</b> encrypts to?	R	K	V	C	B	L	J	
Close to $freq(en) = 111$ .	<b>H</b>	102	163	61	25	148	76	71
Close to $freq(tn) = 7$ .	<b>U</b>	66	71	82	107	31	0	5
Close to $freq(hn) = 1$	<b>G</b>	92	15	31	41	0	5	0
The three highest frequency entries should be a reasonable match for the 181, 179 and 106 (the frequencies of <b>an</b> , <b>in</b> and <b>on</b> .)	<b>R</b>	---	<u>87</u>	0	15	<u>133</u>	<u>143</u>	<u>66</u>
	<b>K</b>	<u>87</u>	---	<u>51</u>	<u>92</u>	10	10	25
	<b>V</b>	10	36	---	0	<u>66</u>	<u>133</u>	<u>56</u>
	<b>C</b>	25	<u>102</u>	<u>66</u>	---	<u>25</u>	<u>133</u>	<u>56</u>
	<b>B</b>	<u>46</u>	<u>61</u>	<u>97</u>	<u>46</u>	---	20	10
	<b>L</b>	36	46	46	15	5	---	0
	<b>J</b>	<u>76</u>	5	36	<u>56</u>	5	0	---

The only good matches in the table above occur in the column for **L**, and in the bottom part of the table they occur in the rows for **R**, **V**, and **C**. So it is very likely that **n** → **L**, and  $\{a, o, i\} \rightarrow \{R, V, C\}$  in some order.

**3) It is likely that  $a \rightarrow R$ ,  $o \rightarrow V$ ,  $i \rightarrow C$ .**

We already know each of **a**, **o**, and **i** encrypts to one of **R**, **V**, or **C**. Consider the following table of single-letter and digram frequencies. Each box contains a letter or digram and its frequency.

a	821	aa	1	ae	0	ea	110	at	104	ta	59	ah	5	ha	114	an	181	na	40
o	767	oo	36	oe	5	eo	33	ot	57	to	115	oh	11	ho	49	on	106	no	60
i	676	ii	1	ie	23	ei	41	it	93	ti	76	ih	6	hi	97	in	179	ni	33

R	854	RR	5	RH	5	HR	102	RU	158	UR	66	RG	5	GR	92	RL	143	LR	36
V	743	VV	0	VH	10	HV	61	VU	41	VR	82	VG	5	GV	31	VL	133	LV	46
C	642	CC	0	CH	31	HC	25	CU	51	CR	107	CG	0	GC	41	CL	133	LC	15

Consider the row for **a**. It matches the row for **R** fairly reasonably in all positions, and very well in most positions. It doesn't match the row for **V** as well due to major variations in the columns for **at** and **ha**. And it doesn't match the row for **C** well at all; note especially the columns for **a**, **ea**, and **ha**.

So we conclude it is likely that **a → R**.

We can try to match the row for **o** with the rows for **V** or **C**, but there is no clear winner.

To separate **i** and **o**, we can make use of the fact that **ou** is a very common digram ( $freq=115$ ), while **iu** is quite rare ( $freq=1$ ). The plaintext letter **u** ( $freq = 291$ ) should encrypt to a ciphertext letter of somewhat comparable frequency — probably a letter between **S** ( $freq = 204$ ) and **W** ( $freq = 397$ ) in the list of ciphertext letters, sorted by frequency. These letters are **S, X, Q, I, E, W**.

Consider this table of digram frequencies. The first letter of the digram is taken from the column at left, the second from the row at top.

	S	X	Q	I	E	W
V	15	102	71	10	82	15
C	36	10	0	25	20	15

Only **VX**, **VQ**, and **VE** are reasonable candidates for **ou** encrypted. In any case, **o → V**. This leaves us with **i → C**.

#### 4) It is likely that **s → K**.

Consider what letter encrypts to **K** ( $freq = 804$ ). All plaintext letters with frequencies higher than **s** ( $freq=617$ ) have been accounted for. Although the frequencies of **s** and **K** are a bit further apart than we would expect for **s→K**, any other letter that could encrypt to **K** would have a frequency of 550 or less — a substantially worse discrepancy. So we assume **s→K**.

At this point, we have deduced the probable encryptions of the eight most common plaintext letter. Here we decrypt the text, using a dot to indicate a letter we do not yet know how to decrypt.

```
a.a.s..e.a...in...st..o.a.o.e.t.hi.h.o...she..i.h.t.on.the.o.i.in.s.o.o....an.e.t.a..s.st.e
.has.an.e.ons.he...ein.the.nite.statesat.theen.o.asi..ea..ission.the.sta...st.et..n.a..s..es.o.e..o...hto.eet.e.se.on.asit.et..ne.ata..a..in.the.i.sts...ess....o..e.tiono..o.eta..an.inte.ste..a..ate.ia.an...a.in..hee.s..o.nasas.ientists.oe.e..i
n.aan.se.e.a..o.en.o..ea..es.ath.e.e.an.io.s...este..a.ata.i.ita..testin..an.ein.tah.at.hin.asthe..o.ethesi.eo.a.a..e..ant.oth..t.e.intoheat.os.he.e.aste.than
..an.a.eo..e.tha..one.e.o.eas.o.etsa.e.a..e.a..so..i.t.i.e.e.to.e...o.the..eatio
no.theso.a.s.ste..i..ion.ea.sa.os.ientistse..e.tthatthesa...es.i...e.ea....esa.o
.tho.the..anets.e.e..o..e.an..ossi...theo.i.inso..i.e..e..in.asne..o.s.is.ositio
n.asnotso.e...e.a seo.thehi.hs.ee....inationo.the.ission.o..hi.hheis..o..a..e.a
na.e..o..o..hee..a.tins.a.es.ste.sthe.o..an.that.esi.ne.an..onst...te.the..o.ean
.its.othe.shi.si.teen.ontha.ohe.at.he.inho..o.asthe.enesis.et..n..o.e.a...in.so
..a..in..a..ti..es..i.t..thesa.etea..n.e.his.i.e.tionan..ase.onsi.i.a.te.hno.o..s
.a..e..into.the..o..n..at..he..e...o..is..on..i..entin..hat..eha..e.esi..ne..a..tan..teste
.hesai...te.e...o..a..a..s..o..iesa..o..t..hat..eha..enottho..hto..sta...st..as..a..n..he..in
..e..a..iten..o..nte..e..the..i..o..ett..o..ea..sa..o..an..ho..e.....athe..e...st..o..itstai..
in..a..a..esha..e..i..eatennis..a...etae..o..e..a..o..o...ess..aseo..s..e...thatisthe..i..htes
t..ate..ia..onea..than..is..no..ntos..ientistsas..o..ens..o..e..as..se..intheat..ttot..a..the..ate..ia....att..en..e..o..i..e..ia..o..e..e..on..ona..ea..in..e..e..ton..t..ate..est..ia...st..i
...eoneo..the..i..st..i..tishs..ientiststo..e..ei..e..sts..a...estoana..sethisisthe..i..stti..esin..ethea..o..o..ission..thatsa...eso..o..ha..e..een..et..ne..o..s..a..etoea..thhesai..t
histo..san..tho..a..a..eo...st..o..i..i..o..a..a..te..s..o..ea..o..t..the..o..ationo..the..so..a..s..ste..than..the..ast..ea..so..te..es..o..eo..se..ation..oso..these..e..tsitisa..eatti..eto
..einto..st..an..e..ie..ethat..o..ets..e..i..e..e..osto..the..ate..onea..than..ossi..theo..an
i..o..o..n..sthat..e..e..ita..to..hee..o..t..t..i..o..i..e..
```

At this point, we could finish the decoding by inspection of the text above. It is easier if we discover the encodings of a few more letters by frequency analysis.

### 5) It is likely that $d \rightarrow W$ , $r \rightarrow B$ , and $I \rightarrow J$ .

Here is a table of useful letter and digram frequencies for **r**, **d**, and **I**, followed by the corresponding frequencies of ciphertext letters/digrams to which they might reasonably encrypt (based on character frequencies).

$\lambda$	Plaintext estimated frequency of $\lambda$ and digrams involving $\lambda$																	
	$\lambda$	$\lambda\lambda$	$\lambda e$	$e\lambda$	$\lambda t$	$t\lambda$	$\lambda h$	$h\lambda$	$\lambda a$	$a\lambda$	$\lambda n$	$n\lambda$	$\lambda o$	$o\lambda$	$\lambda i$	$i\lambda$	$\lambda s$	$s\lambda$
<b>r</b>	550	14	133	169	42	28	12	8	50	75	16	3	55	84	50	27	37	4
<b>d</b>	479	13	57	126	56	7	25	1	48	52	16	146	41	18	50	33	35	7
<b>I</b>	393	56	64	55	15	17	4	2	40	57	2	9	41	26	47	37	11	13

$\Phi$	Ciphertext frequency of $\Phi$ and digrams involving $\Phi$																	
	$\Phi$	$\Phi\Phi$	$\Phi H$	$H\Phi$	$\Phi U$	$U\Phi$	$\Phi G$	$G\Phi$	$\Phi R$	$R\Phi$	$\Phi L$	$L\Phi$	$\Phi V$	$V\Phi$	$\Phi C$	$C\Phi$	$\Phi K$	$K\Phi$
<b>B</b>	616	25	102	148	51	31	5	0	46	133	20	5	97	66	46	25	61	10
<b>J</b>	484	76	87	71	15	5	0	0	76	66	0	0	36	56	56	56	5	25
<b>W</b>	397	20	36	138	25	5	10	0	41	20	0	107	31	15	61	15	20	15
<b>E</b>	316	25	61	46	10	20	5	0	61	51	5	10	31	82	41	20	10	15
<b>I</b>	280	10	31	66	31	15	41	5	25	31	0	25	87	10	20	25	0	41
<b>Q</b>	260	0	15	5	31	20	0	5	0	10	31	5	0	71	10	0	82	20

**d** might reasonably encrypt to **B**, **J**, **W**, or possibly **E**. But only **LW** ( $freq=107$ ) comes anywhere close to matching the high frequency of **nd** ( $freq=146$ ). Noting also that the entire row for **W** matches that for **d** fairly well, we assume **d**→**W**.

**r** might reasonably encrypt to **B** or **J**. **BH** and **HB** provides a significantly better match for the high-frequency digrams **re** and **er**, than do **JH** and **HJ**. In fact, in 14 columns of the table, row **r** more closely matches row **B** than row **J**, in two col-

umns it matches row **J** more closely, and in two columns there is a tie. (In many rows, the differences are not significant by themselves.). So we assume **r**→**B**.

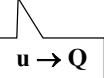
Suppose **I** does not encrypt to **J**? The next-highest frequency plaintext letter available is **u**. It seems unlikely that **u** ( $freq=291$ ) would encrypt to **J** ( $freq=484$ ).

We have deduced the probable encryptions of the eleven most common plaintext letters. Here is our text with these eleven letters decrypted.

atapsleiarriainsd~~st~~roeaioet~~hi~~oldshedlightontheorisins~~xoqr~~planetartsiste  
haslandedons~~hed~~leinthe~~q~~itedstatesattheendo~~xasi~~earissionthestard~~stre~~~~rn~~  
~~ap~~~~leslo~~~~ed~~~~xro~~~~EEP~~~~hto~~~~xeet~~~~perse~~~~ondasitret~~~~grnedata~~~~EE~~~~arp~~~~in~~~~the~~~~ir~~~~sts~~~~QII~~~~ess~~~~XQI~~~~olle~~  
~~ition~~~~XI~~~~tar~~~~andinterstellar~~~~p~~~~aterial~~~~anddr~~~~azin~~~~S~~~~heers~~~~ro~~~~nasas~~~~I~~~~ientists~~~~co~~~~elli~~  
~~n~~~~sa~~~~andse~~~~meraldo~~~~en~~~~iollea~~~~S~~~~ess~~~~atheredan~~~~Ni~~~~o~~~~sl~~~~T~~~~esterda~~~~ata~~~~E~~~~ilitar~~~~testin~~~~r~~~~an~~~~ein~~  
~~ta~~~~z~~~~ati~~~~hin~~~~s~~~~as~~~~the~~~~Pro~~~~thesi~~~~eo~~~~xal~~~~arse~~~~plant~~~~P~~~~oth~~~~rtled~~~~into~~~~heat~~~~los~~~~ph~~~~ex~~~~aster~~~~than~~  
~~tean~~~~E~~~~adeo~~~~F~~~~O~~~~E~~~~ithad~~~~done~~~~F~~~~oreas~~~~Io~~~~etsare~~~~large~~~~forallso~~~~dirt~~~~ti~~~~ele~~~~to~~~~mer~~~~xro~~~~the~~~~I~~~~ratio~~  
~~no~~~~x~~~~thesolars~~~~st~~~~E~~~~FF~~~~illion~~~~years~~~~as~~~~os~~~~ientist~~~~se~~~~N~~~~PE~~~~tt~~~~that~~~~thes~~~~a~~~~EP~~~~les~~~~zill~~~~re~~~~real~~~~I~~~~Q~~~~esa~~  
~~po~~~~Q~~~~tho~~~~the~~~~planets~~~~ze~~~~xor~~~~ed~~~~and~~~~Possi~~~~Flt~~~~theori~~~~is~~~~in~~~~xo~~~~li~~~~xe~~~~Mellin~~~~Asner~~~~M~~~~Q~~~~sd~~~~ispositio~~  
~~n~~~~z~~~~asnot~~~~sole~~~~el~~~~T~~~~Fe~~~~ia~~~~Q~~~~seo~~~~x~~~~thehi~~~~Shs~~~~eed~~~~IQ~~~~l~~~~ination~~~~o~~~~x~~~~the~~~~E~~~~ission~~~~xor~~~~zhi~~~~heis~~~~pro~~~~rae~~~~ee~~  
~~a~~~~na~~~~er~~~~xor~~~~lo~~~~ID~~~~heed~~~~art~~~~in~~~~spa~~~~es~~~~T~~~~st~~~~e~~~~the~~~~o~~~~EP~~~~an~~~~tha~~~~des~~~~is~~~~ned~~~~and~~~~on~~~~str~~~~Q~~~~ted~~~~the~~~~pro~~~~ean~~  
~~dits~~~~o~~~~ther~~~~ship~~~~s~~~~int~~~~een~~~~on~~~~tha~~~~s~~~~ohe~~~~zat~~~~I~~~~he~~~~in~~~~horror~~~~as~~~~the~~~~enes~~~~is~~~~ret~~~~Q~~~~rn~~~~ro~~~~Fe~~~~arr~~~~in~~~~so~~  
~~lar~~~~z~~~~ind~~~~parti~~~~iles~~~~F~~~~Q~~~~ilt~~~~FT~~~~the~~~~sa~~~~e~~~~tea~~~~EQ~~~~nder~~~~his~~~~dire~~~~tion~~~~and~~~~ased~~~~on~~~~si~~~~elar~~~~t~~~~hnolo~~~~ST~~~~sl~~  
~~a~~~~EE~~~~ed~~~~into~~~~the~~~~s~~~~ro~~~~ndat~~~~E~~~~he~~~~mer~~~~TF~~~~od~~~~is~~~~on~~~~X~~~~ident~~~~in~~~~zha~~~~eha~~~~med~~~~esi~~~~sned~~~~and~~~~F~~~~ilt~~~~and~~~~tested~~  
~~hesaid~~~~F~~~~Q~~~~te~~~~Mer~~~~TF~~~~od~~~~T~~~~al~~~~z~~~~sorriesa~~~~po~~~~Q~~~~z~~~~hat~~~~zha~~~~menot~~~~tho~~~~Q~~~~sho~~~~stard~~~~Q~~~~st~~~~z~~~~asla~~~~Q~~~~n~~~~ihedin~~  
~~x~~~~er~~~~Q~~~~ar~~~~T~~~~iten~~~~I~~~~Q~~~~ntered~~~~the~~~~zild~~~~Io~~~~ett~~~~z~~~~o~~~~ears~~~~s~~~~and~~~~hope~~~~EQ~~~~ll~~~~TS~~~~athered~~~~d~~~~Q~~~~st~~~~xro~~~~it~~~~tail~~~~Q~~  
~~in~~~~SP~~~~addle~~~~shape~~~~lid~~~~de~~~~aten~~~~nisra~~~~IA~~~~eta~~~~ero~~~~sel~~~~alo~~~~Q~~~~ress~~~~s~~~~ase~~~~Q~~~~ell~~~~tha~~~~is~~~~the~~~~light~~  
~~ter~~~~rial~~~~alone~~~~arth~~~~and~~~~dis~~~~no~~~~z~~~~tos~~~~ient~~~~ist~~~~s~~~~xro~~~~en~~~~se~~~~ez~~~~as~~~~sed~~~~in~~~~the~~~~atte~~~~EP~~~~tt~~~~trap~~~~the~~  
~~ater~~~~ial~~~~dr~~~~att~~~~sen~~~~sex~~~~ro~~~~ei~~~~per~~~~iol~~~~le~~~~Se~~~~lon~~~~don~~~~ale~~~~ad~~~~in~~~~SEND~~~~ER~~~~one~~~~tr~~~~at~~~~re~~~~trial~~~~Q~~~~z~~  
~~ill~~~~Pe~~~~one~~~~Q~~~~the~~~~x~~~~irst~~~~Fr~~~~iti~~~~sh~~~~is~~~~ent~~~~ist~~~~store~~~~ie~~~~imed~~~~Q~~~~sts~~~~a~~~~EP~~~~lest~~~~o~~~~anal~~~~set~~~~is~~~~the~~~~x~~~~irst~~~~ti~~  
~~E~~~~es~~~~in~~~~the~~~~a~~~~pollo~~~~E~~~~ission~~~~st~~~~tha~~~~tsa~~~~EP~~~~les~~~~o~~~~ro~~~~ID~~~~ha~~~~e~~~~een~~~~ret~~~~Q~~~~rn~~~~ed~~~~xro~~~~SP~~~~ai~~~~to~~~~ear~~~~hesaid~~  
~~hist~~~~ho~~~~Q~~~~s~~~~and~~~~tho~~~~x~~~~AS~~~~ra~~~~EE~~~~eo~~~~d~~~~st~~~~xro~~~~E~~~~ild~~~~zill~~~~pro~~~~f~~~~lt~~~~tell~~~~Q~~~~s~~~~orea~~~~fo~~~~Q~~~~the~~~~or~~~~ation~~~~o~~~~the~~  
~~solars~~~~T~~~~st~~~~E~~~~than~~~~the~~~~past~~~~tear~~~~so~~~~X~~~~teles~~~~Io~~~~eo~~~~F~~

Consider the second line. I have inserted a little space at some likely word breaks.

E has landed on schedule in the United States at the end of a six year mission the Stardust return

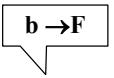
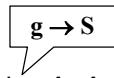
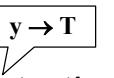
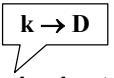


E has landed on schedule in the United States at the end of a six year mission the Stardust return



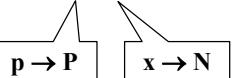
m has landed on schedule in the United States at the end of a six year mission the Stardust return

Now let us look at the 11<sup>th</sup> and 12<sup>th</sup> lines, with the characters above decoded.



na s er for loc D he ed mart in s pac es T stem s the com pa ny th at de si gn ed and co nstruc ted the pro be an

dits mo the r shi ps i n te en mon th s a go he wa tched in horro r as the ge ne sis re turn pro be car ry ing so

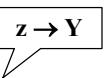
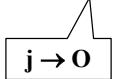


n a g e for lock he ed mar tin s pac es sys tem s the com pa ny th at de si gn ed and co nstruc ted the pro be a n

dits mo the r shi ps si xteen mon th s a go he wa tched in horro r as the ge ne sis re turn pro be car ry ing so

So far we have decrypted 22 of the 26 characters; these lines will yield three more.

y man made ob ject ha ddone before as comets are large balls of dirty ice left over from the creatio



tah watching as the prob e the si ve of a large plant po thurt led into the atmos phere faster than an

The final plain text character q must encrypt to the final ciphertext character A. So the encryption key is the permutation

( a b c d e f g h i j k l m n o p q r s t u v w x y z )  
( R F I W H X S G C O D J E L V P A B K G Q M Z N T N )

and our plaintext was

a capsule carrying dust from a comet which could shed light on the origins of our planetary system  
m has landed on schedule in the United States at the end of a six year mission the Stardust return  
ncapsules slowed from mph to feet per second as it returned data marking the first successful collection of cometary and interstellar material and drawing cheers from NASA scientists Joe Velli  
ng and several dozen colleagues gathered anxiously yesterday at a military testing range in Utah watching as the probe the size of a large plant pothurtled into the atmosphere faster than an  
uman-made object had done before as comets are large balls of dirty ice left over from the creation  
of the solar system billion years ago. Scientists expect that the samples will reveal clues about how the planets were formed and possibly the origins of life. In the meantime, the samples will reveal clues about the formation of the solar system.  
It was not solely because of the high speed culmination of the mission for which the program manager for Lockheed Martin Space Systems, the company that designed and constructed the probe, was not alone. The probe carried six wind particles built by the same team under his direction and based on similar technology. It was intended to ground at the speed of light, but it was not always confident in what it had been designed and built and tested. He said but everybody always worries about what we have thought of Stardust was launched in February and it encountered the wild comet two years ago and hopefully gathered dust from its tail. It was a paddle-shaped like a tennis racket and a colourless gaseous jelly that is the lightest material known to scientists as frozen smoke was used in the attempt to trap the probe. It was a matter of time before the first British scientist to receive dust samples to analyze this is the first time since the Apollo 11 mission that samples of rock have been returned from space to Earth. He said it took thousands of grams of dust from wild will probably tell us more about the formation of the solar system than the past year of telescope observations.

It is interesting to compare the frequencies and ranks of the ciphertext letters (and of the actual plaintext letters) with those predicted from our “typical” English language text data.

Letter		Frequency of Letter		Rank of Letter	
typical plain text	cipher text	typical plain text	cipher text	typical plain text	cipher text
e	H	1237	1253	1	1
t	U	921	1013	2	2
a	R	821	845	3	3
o	V	767	743	4	5
n	L	705	580	5	8
i	C	676	642	6	6
h	G	645	479	7	10
s	K	617	804	8	4
r	B	550	616	9	7
d	W	479	397	10	11
l	J	393	484	11	9
u	Q	291	260	12	14
w	Z	254	137	13	20
m	E	254	316	14	12
c	I	230	280	15	13
f	X	225	229	16	15
g	S	208	204	17	16
y	T	195	178	18	18
p	P	163	199	19	17
b	F	150	153	20	19
k	D	87	31	21	22
v	M	87	87	22	21
j	O	18	20	23	24
x	N	13	31	24	22
q	A	9	5	25	26
z	Y	6	15	26	25