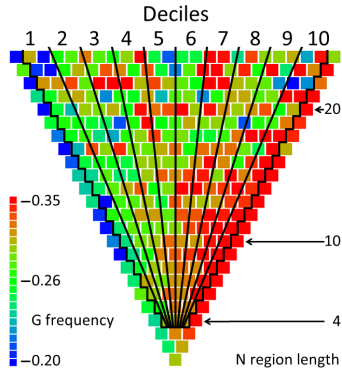


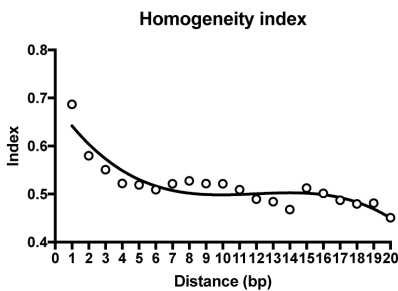
Supplementary Figures and tables

Supplementary Figure 1



Supplementary Figure 1 Heat map illustrating the nucleotide usage (in the presented case G) across N regions of varying lengths, from low usage (blue) to high usage (red). The black lines indicate the partitioning into deciles after one nucleotide has been removed from each end of the N regions (Stair-case line).

Supplementary Figure 2



Supplementary Figure 2 Homogeneity index as defined by Kepler et al. (1) calculated for 6,924 N2 regions from NP unmutated rearrangements with trimmed flanking gene segments. High values above 1/2 may be caused by nucleotide stacking. With increasing distance between the nucleotide pairs the single-strand polymerization hypothesis would predict consistent values above 1/2 while the concatenation hypothesis would predict declining values ending below 1/2 for large values of distance. The sensitivity is much larger for the single-strand polymerization hypothesis than for the concatenation hypothesis, because in the latter case, the nucleotide pairs may be synthesized in the same strand or in opposite strands which will tend to yield values close to 1/2. The finding argues against the single-strand polymerization hypothesis and does not argue against the concatenation hypothesis.

$$\text{Homogeneity index} = \frac{\text{fn}(G..G) + \text{fn}(C..C)}{\text{fn}(G..C) + \text{fn}(C..G) + \text{fn}(G..G) + \text{fn}(C..C)}$$

where $f_n(N..N)$ is the frequency of nucleotide pairs of the given kind in the N region separated by $n-1$ nucleotides (1).

Supplementary Table 1.

Supplementary Table 1.		
7-mer	6-mer	5-mer
TGAGAAA	CGAAA	CGGAT*
CTAGGGA	GAGAGA	GAAAG
CTAGAGA	CGGATA	AGATA
CCAGAGA	CAGAGA	TGAGA
CGGCAGA	AAAGAT	ACGGA
CGGATAC	GAGAGG	CACAC
CGAGATA	GAGATA	AAGAT
CGAGACA	CAGACC	AGAGA
CGAGAGG	GAGACA	GAGAT
CGAGAGA	GAGAAA	CAAAA
CGAGAAA	TGAGAA	AGGGA
CGAAAAGA	CGAGAA	GAGAG
CACGGAT	TAGGGA	CAGAC
CACACAG	CGAGAC	ACACA
CACAGAC	CGAGAG	AAAAG
CAAGAGA	CCAGAG	CCAGA
CAAAAAGA	CAAAAAG	TAGAG
ACGGATA	AAAAGA	CACGG
ACACAGA	TAGAGA	TAGGG
ACAGACC	GGATAC	AGAGG
AAAGATA	CTAGAG	CAAGA
AAAAGAT	GAAAAGA	AGACC
	CTAGGG	GATAC
	CACGGA	CTAGG
	CGGCAG	AAGAG
	CACAGA	CGGCA
	AAGAGA	CTAGA
	AAGATA	CACAG
	CGAGAT	GGATA
	CACACA	GAGAC
	CAAGAG	CGAAA
	ACGGAT	AGACA
	ACAGAC	CGAGA
	GGCAGA	GAGAA
	ACACAG	CAGAG
		AAAGA
		AGAAA
		GGCAG
		GCAGA
		ACAGA

*) only motif found more often in N1 than N2

References

1. Kepler, T. B., M. Borrero, B. Ruggiero, S. K. McCray, and S. H. Clarke. 1996. Interdependence of N nucleotide addition and recombination site choice in V(D)J rearrangement. *J. Immunol.* 157: 4451–4457.