Pairix Index Binary Spec

Soo Lee (duplexa@gmail.com)

Category			Field	Description	Type	Value
magic			magic	Magic string	char[8]	PX2.004\1 old magic numbers : PX2.003\1 for pairix 0.3.4/0.3.5 PX2.002\1 for pairix <=0.3.3
n_ref			n_ref	# sequences	int32_t	
linecount			linecount	# lines	uint64_t / int32_t	unit64_t or latest magic number, int32_t for old magic numbers
conf			format	Format (0: generic; 1: SAM; 2: VCF; 3: PAIRS, 4: MERGED_NODUPS, 5: OLD_MERGED_NODUPS)	int32_t	
			col_seq	Column for the sequence name	int32_t	
			col_beg	Column for the start of a region	int32_t	
			col_end	Column for the end of a region	int32_t	
			col_seq2	Column for the 2nd sequence name	int32_t	
			col_beg2	Column for the start of the 2nd region	int32_t	
			col_end2	Column for the end of the 2nd region	int32_t	
			delimiter	column delimiter	char	default '\t' except MERGED_NODUPS and OLD_MERGED_NODUPS (' ')
			region_split_char	characterthat splits two regions in a 2D query and two sequence names in the index	char	default '|'
				2 bytes padded		
			meta	Leading character for comment lines	int32_t	default '#', except SAM format ('@')
			skip	# lines to skip at the beginning	int32_t	default 0
target names			l_nm	Length of concatenated sequence pair names	int32_t	
			names	Concatenated sequence pair names, each zero terminated	char[l nm]	
index			*List of indices (n=n_ref)*			
	binning index		n_bin	# distinct bins (for the binning index)	int32_t	
			List of distinct bins (n=n_bin)			
			bin	Distinct bin number	uint32 t	
			n_chunk	# chunks	int32_t	
			List of chunks (n=n_chunk)			
		cnk_beg	Virtual file offset of the start of the chunk	uint64 t	
		cnk_end	Virtual file offset of the end of the chunk	uint64 t	
	linear index		n_intv	# 16kb intervals (for the linear index)	int32 t	
			List of distinct intervals (n=n_intv)			
		ioff	File offset of the first record in the interval	uint64 t	