TCB No. 005-2009 April 2009

Technical Bulletin

Genome Sequencer FLX System

Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Extended MID Set

Summary

This bulletin describes the use of a set of up to 151 Multiplex Identifiers (MID) with the GS FLX Titanium chemistry. This document is intended to be an extension of Technical Bulletin 09004: *Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Basic MID Set*. Libraries prepared with these Adaptors may be multiplexed in emulsion PCR (emPCR) in order to enable sample identification following sequencing on the Genome Sequencer FLX Instrument. This bulletin enables much deeper multiplexing for users who wish to employ the 10-base barcode strategy provided by the MID Adaptors.

Roche Diagnostics GmbH
Roche Applied Science
68298 Mannheim, Germany

Roche Diagnostics Corp.
Roche Applied Science
Indianapolis, IN, 46250 USA

Note: This bulletin assumes that the user is familiar with the contents of *TCB-09004: Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Basic MID Set.* Please consult this document for specific instructions regarding the preparation of MID Adaptors as well as for the handling of libraries made with the same.

Introduction

The information contained in this document is provided to enable users of the GS FLX Titanium sequencing chemistry to employ Multiplex Identifier (MID)-containing adaptors for General (*e.g.* Shotgun) library preparation. Please note that the *GS FLX Titanium General Library Preparation Method Manual* and the GS FLX Titanium General Library Preparation Kit can be used to prepare a library without reference to the information contained in this bulletin. If you are preparing standard, non-MID libraries, you do not need this document.

This document describes the preparation and use of up 151 Multiplex Identifier (MID) Adaptors. This extends the Basic MID Set of ten Adaptors for users requiring greater multiplexing capabilities with the GS FLX Titanium sequencing chemistry. These MID Adaptors may be used as a replacement for the Adaptors provided in the GS FLX Titanium General Library Preparation Kit. These Adaptors include a 10-nucleotide sequence tag on Adaptor A which is unique for each MID. When different libraries are prepared with different MIDs, they can be amplified by emPCR and sequenced together, in a multiplex fashion; the sequencing reads can be deconvoluted by the data analysis software after the sequencing Run, such that the reads from each of the pooled libraries are identified by their MID tag and correctly assigned.

Extended Multiplex Identifier Set Design

A robust set of ten decamer Multiplex Identifier (MID) sequences (Basic MID Set, described previously in *TCB-09004: Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Basic MID Set.*) was designed to facilitate library multiplexing in the 454 Sequencing system. A length of ten bases is sufficient to ascertain that, for the large number of reads involved and the design parameters considered, the chances of mis-assigning reads is extremely low. By relaxing some of the design requirements, an even greater number of 10-base MIDs can be devised for library multiplexing purposes. Relaxing the requirements for shotgun libraries is not expected to result in a significant loss of reads, because sequencing is highly accurate just beyond the key sequence, at the beginning of sequencing reads. This larger set of MID Adaptors is know as the Extended MID Set:

- The Extended MID Set is listed in Table 1. Each MID sequence is at least 4 changes (insertion, deletion, substitution) away from the other members of the Extended MID Set. This means that for any of these MIDs, it is possible to either detect up to 2 errors and correct 1 error, or alternatively, detect 3 errors and correct none.

- The first ten sequences in the list (highlighted in bold text in Table 1) represent the Basic Set MIDs. To this set have been added 141 MIDs to comprise the final set of 151.

- The Extended Set MIDs are sorted according to the number of reagent flows needed to sequence each, with lower number meaning fewer flows. As a result, the lower numbered entries of the Extended Set MIDs should be preferred over the higher numbered Adaptors, because they can be sequenced using fewer reagent flows thereby maximizing the number of remaining flows for sequencing the library fragment.

Table 1. 10-base Extended Multiplex Identifier (MID) Set Sequences

MID-1	**ACGAGTGCGT**		MID-40	TACGCTGTCT
MID-2	**ACGCTCGACA**		MID-41	TAGTGTAGAT
MID-3	**AGACGCACTC**		MID-42	TCGATCACGT
MID-4	**AGCACTGTAG**		MID-43	TCGCACTAGT
MID-5	**ATCAGACACG**		MID-44	TCTAGCGACT
MID-6	**ATATCGCGAG**		MID-45	TCTATACTAT
MID-7	**CGTGTCTCTA**		MID-46	TGACGTATGT
MID-8	**CTCGCGTGTC**		MID-47	TGTGAGTAGT
MID-10	**TCTCTATGCG**		MID-48	ACAGTATATA
MID-11	**TGATACGTCT**		MID-49	ACGCGATCGA
MID-13	CATAGTAGTG		MID-50	ACTAGCAGTA
MID-14	CGAGAGATAC		MID-51	AGCTCACGTA
MID-15	ATACGACGTA		MID-52	AGTATACATA
MID-16	TCACGTACTA		MID-53	AGTCGAGAGA
MID-17	CGTCTAGTAC		MID-54	AGTGCTACGA
MID-18	TCTACGTAGC		MID-55	CGATCGTATA
MID-19	TGTACTACTC		MID-56	CGCAGTACGA
MID-20	ACGACTACAG		MID-57	CGCGTATACA
MID-21	CGTAGACTAG		MID-58	CGTACAGTCA
MID-22	TACGAGTATG		MID-59	CGTACTCAGA
MID-23	TACTCTCGTG		MID-60	CTACGCTCTA
MID-24	TAGAGACGAG		MID-61	CTATAGCGTA
MID-25	TCGTCGCTCG		MID-62	TACGTCATCA
MID-26	ACATACGCGT		MID-63	TAGTCGCATA
MID-27	ACGCGAGTAT		MID-64	TATATATACA
MID-28	ACTACTATGT		MID-65	TATGCTAGTA
MID-29	ACTGTACAGT		MID-66	TCACGCGAGA
MID-30	AGACTATACT		MID-67	TCGATAGTGA
MID-31	AGCGTCGTCT		MID-68	TCGCTGCGTA
MID-32	AGTACGCTAT		MID-69	TCTGACGTCA
MID-33	ATAGAGTACT		MID-70	TGAGTCAGTA
MID-34	CACGCTACGT		MID-71	TGTAGTGTGA
MID-35	CAGTAGACGT		MID-72	TGTCACACGA
MID-36	CGACGTGACT		MID-73	TGTCGTCGCA
MID-37	TACACACACT		MID-74	ACACATACGC
MID-38	TACACGTGAT		MID-75	ACAGTCGTGC
MID-39	TACAGATCGT		MID-76	ACATGACGAC

MID-77	ACGACAGCTC
MID-78	ACGTCTCATC
MID-79	ACTCATCTAC
MID-80	ACTCGCGCAC
MID-81	AGAGCGTCAC
MID-82	AGCGACTAGC
MID-83	AGTAGTGATC
MID-84	AGTGACACAC
MID-85	AGTGTATGTC
MID-86	ATAGATAGAC
MID-87	ATATAGTCGC
MID-88	ATCTACTGAC
MID-89	CACGTAGATC
MID-90	CACGTGTCGC
MID-91	CATACTCTAC
MID-92	CGACACTATC
MID-93	CGAGACGCGC
MID-94	CGTATGCGAC
MID-95	CGTCGATCTC
MID-96	CTACGACTGC
MID-97	CTAGTCACTC
MID-98	CTCTACGCTC
MID-99	CTGTACATAC
MID-100	TAGACTGCAC
MID-101	TAGCGCGCGC
MID-102	TAGCTCTATC
MID-103	TATAGACATC
MID-104	TATGATACGC
MID-105	TCACTCATAC
MID-106	TCATCGAGTC
MID-107	TCGAGCTCTC
MID-108	TCGCAGACAC
MID-109	TCTGTCTCGC
MID-110	TGAGTGACGC
MID-111	TGATGTGTAC
MID-112	TGCTATAGAC
MID-113	TGCTCGCTAC
MID-114	ACGTGCAGCG
MID-115	ACTCACAGAG

MID-116	AGACTCAGCG
MID-117	AGAGAGTGTG
MID-118	AGCTATCGCG
MID-119	AGTCTGACTG
MID-120	AGTGAGCTCG
MID-121	ATAGCTCTCG
MID-122	ATCACGTGCG
MID-123	ATCGTAGCAG
MID-124	ATCGTCTGTG
MID-125	ATGTACGATG
MID-126	ATGTGTCTAG
MID-127	CACACGATAG
MID-128	CACTCGCACG
MID-129	CAGACGTCTG
MID-130	CAGTACTGCG
MID-131	CGACAGCGAG
MID-132	CGATCTGTCG
MID-133	CGCGTGCTAG
MID-134	CGCTCGAGTG
MID-135	CGTGATGACG
MID-136	CTATGTACAG
MID-137	CTCGATATAG
MID-138	CTCGCACGCG
MID-139	CTGCGTCACG
MID-140	CTGTGCGTCG
MID-141	TAGCATACTG
MID-142	TATACATGTG
MID-143	TATCACTCAG
MID-144	TATCTGATAG
MID-145	TCGTGACATG
MID-146	TCTGATCGAG
MID-147	TGACATCTCG
MID-148	TGAGCTAGAG
MID-149	TGATAGAGCG
MID-150	TGCGTGTGCG
MID-151	TGCTAGTCAG
MID-152	TGTATCACAG
MID-153	TGTGCGCGTG

Note: While the Extended MID Set has been designed to the best of our ability using all current knowledge, not all sequences have yet been thoroughly tested in library construction. It is possible that one or more MIDs on the list may not perform as expected. Our experience to date with 10-base MIDs has shown that approximately one in ten sequences performs below expectations, because of unexpected dimerization/ligation events or unanticipated PCR amplification artifacts. Please note that MID-9 and MID-12 are only used for Standard Series, not Titanium Series methods, and are intentionally excluded herein.

Obtaining and Preparing MID Adaptors

1. For each different MID desired, you must obtain the required oligonucleotides and prepare a new Adaptors mix. Each MID Adaptors mix is comprised of an Adaptor A and an Adaptor B. Each adaptor is comprised of two oligonucleotides that are annealed in an equimolar ratio and these adaptors are then combined to make an MID Adaptors mix.

2. All oligonucleotides should be obtained according to the processes and specifications outlined in the document *Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Basic MID Set.* Importantly, oligonucleotide synthesis specifications for the Extended MID Set are identical to those for the Basic MID Set:

    a. Each oligonucleotide should have phosphorothioate modifications in both the first four and last four bases of the oligomers.
    b. The Adaptor B long oligonucleotide 'Ti-MID-B' must be synthesized with a 5-prime biotin-TEG moiety.
    c. All oligonucleotides must be purified using HPLC.

3. The full sequences of the 306 oligonucleotides that would be required to synthesize all possible Extended MID Set Adaptors A are not provided in this document. However, one can easily design and synthesize the Adaptor A for the particular MID(s) of interest as follows:

    a. Consult the Appendix for a graphical depiction of the structure of the MID-1 Adaptor A (as well as the common Adaptor B which is used with all MID Adaptors mixes).
    b. The highlighted portion of the 'Ti-MID1-A' and 'Ti-MID1-Aprime' oligonucleotides indicates the region of each containing the 10-base MID sequence.
    c. Replace the highlighted portion of the "A" oligonucleotide with the 10-base MID sequence from Table 1 corresponding to the MID of interest. For example, for MID-13:
        1. The 10-base sequence for MID-13 from Table 1 is: CATAGTAGTG
        2. Therefore, the oligonucleotide 'Ti-MID13-A' sequence would be:

        5′-C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGCATAGT*A*G*T*G-3′

    d. Replace the highlighted portion of the "Aprime" oligonucleotide with the reverse complement of the 10-base MID barcode from Table 1 corresponding to the MID of interest. Continuing the example of MID-13:
        1. The 10-base sequence for MID-13 is: CATAGTAGTG and the reverse complement of the same is: CACTACTATG.
        2. Therefore, the oligonucleotide 'Ti-MID13-Aprime' sequence would be :

        5′-C*A*C*T*A*CTATGCTGAGTCG*G*A*G*A-3′

e. It is critical to verify your design by examining the structure of the adaptor that would result from annealing of the two oligonucleotides (including verification of proper Watson-Crick complementary base pairing) as shown in the Appendix for MID-1. Note that you must reverse the left-to-right sequence orientation of the 'Aprime' adaptor to the 3-prime to 5'prime direction in this exercise:

MID-13 Adaptor A:

```
5′-CCATCTCATCCCTGCGTGTCTCCGACTCAGCATAGTAGTG-3′  Ti-MID13-A
                  3′-AGAGGCTGAGTCGTATCATCAC-5′  Ti-MID13-Aprime
```

4. Once ordered and received, oligonucleotides should be annealed and prepared for the Extended MID Set Adaptors mixes according to the procedures given in the document *Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Basic MID Set.*

5. Consult the document *Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Basic MID Set* for instructions on library preparation, emPCR and sequencing.

Appendix

Example: MID Oligonucleotide Sequences for Ordering

OLIGO NAME	OLIGO SEQUENCE (5-prime to 3-prime orientation)
Ti-MID1-A	C*C*A*T*CTCATCCCTGCGTGTCTCCGACTCAGACGAGT*G*C*G*T
Ti-MID1-Aprime	A*C*G*C*ACTCGTCTGAGTCG*G*A*G*A
Ti-MID-B	/5BioTEG/C*C*T*A*TCCCCTGTGTGCCTTGGCAGTC*T*C*A*G
Ti-MID-Bprime	C*T*G*A*GACT*G*C*C*A

- Phosphorothioate bonds are indicated with an asterisk (*)
- A 5'-biotin-TEG moiety is indicated by '/5BioTEG/'
- Inverse (white on black) text denotes the portion of each nucleotide containing the 10-base MID sequence. Note that the highlighted sequence in the "Aprime" oligonucleotide is the reverse complement of that in the associated "A" oligonucleotide.

Purification: All oligonucleotides must be ordered with HPLC purification and with the modifications (i.e. phosphorothioate bonds and 5'-biotin-TEG) shown.

Examples of Annealed MID Adaptors

MID-1 Adaptor A:

```
5′-CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCGT-3′  Ti-MID1-A
                    3′-AGAGGCTGAGTCTGCTCACGCA-5′  Ti-MID1-Aprime
```

MID Adaptor B (Common):

```
5′Biotin-TEG-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3′  Ti-MID-B
                              3′- ACCGTCAGAGTC-5′  Ti-MID-Bprime
```

- Phosphorothioate-modified bases are not shown in this figure for ease of sequence alignment
- Sequencing Key is indicated in GREEN. 10 bp MID sequence is shown in YELLOW.