

DNA marker와 수산양식에 활용(I)

한 현 섭

국립수산과학원 생명공학연구단

1. 서론

DNA marker의 개발로 양식 유전학의 연구 방법은 크게 바뀌었다. 1970년대 처음으로 isozyme이 광범위하게 이용되기 시작한 이후 분자유전학의 비약적인 발전으로, 휴먼 게놈프로젝트 및 기타 유사한 새로운 사업에서 이미 증명된 것처럼, 게놈학은 확고한 위치를 갖게 되었다. 게놈학은 넓은 의미에서 유전자 및 유전자의 기능에 대한 연구로서 정의되고, 건강이나 식품안전성에서부터 임신, 출산과 법 집행 등 다양한 분야에 빠르게 영향을 미치고 있다. 게놈학이 출현하게 된 계기는 DNA marker 기술의 발전이었다. 이들 기술의 발전으로 다양한 유전적 marker가 개발되었고, 이들 marker로는 isozyme, mtDNA, RFLP, RAPD, AFLP, microsatellite, SNP, 그 외 EST가 있고, 다양한 양식 연구에서 널리 이용될 가능성이 있다. 이 총설에서 여러 종류의 marker의 과학적 원리, 잠재력, 제반사항, 장단점을 살펴보고, 더불어 유전적 다양성과 근친도의 평가, 친자 감정, 종 및 계통 동정, 잡종화, 그리고 linkage map을 통한 marker 활용, QTL 동정 등의 활용 사례를 살펴본다. 또한, DNA marker 기술에 대해 체계적으로 소개하고 양식 종사자들에게 적절한 활용 사례를 설명하고자 한다. 그리고 DNA marker 기술의 미래 동향을 비롯해서 휴먼 게놈프로젝트의 성과,

그리고 수산생물 게놈학 및 수산생물 유전학의 발전을 위해 Zebra fish와 같은 모델어류에서 얻은 연구 성과의 활용법에 대해서도 고찰한다.

2. 분자 marker의 종류와 원리

정상적인 세포작용 또는 환경과의 상호작용으로 인해, 모든 생물에서 돌연변이가 일어나는데 이 돌연변이가 유전적 다형성을 유발한다. 선발 및 유전적 부동을 포함하여 개체간, 종간 그리고 더 높은 목 단위 분류 집단내에서도 유전적 변이는 일어난다. 이런 유전적 변이를 유전학자가 이용하려면, 인식할 수 있는 표현형 변이나 분자생화학적 기술을 통해 구분할 수 있는 유전적 돌연변이라도, 이 변이는 (1)유전되어야 하고 (2)연구자가 이를 식별할 수 있어야 한다. DNA 수준에서의 유전적 변이는 일반적으로 단일 뉴클레오타이드 다형성(SNPs)이라고 하는 염기치환 - 한 유전자좌 내에 뉴클레오타이드 배열의 삽입 또는 삭제 - 한 유전자좌내에 DNA 단편의 역위, 한 유전자좌 주변 DNA 단편의 재배열 등이 있다. 이런 DNA 수준에서 유전적 변이는 오랜 기간의 진화과정에서 축적되어, 각각의 돌연변이 형태는 종마다 다양하게 존재하고, 돌연변이의 종류와 정도는 한 종내 유전적 변이를 결정한다. DNA marker 기술은 이런 돌연변이를 찾는데 이용할 수 있다. 변화

가 큰 제거와 삽입은 제한효소에 의해 절단되어 생성되는 DNA 단편의 크기를 변화시키는데, 이런 DNA 변화는 가장 검출하기 쉬운 돌연변이의 하나이며, 주로 agarose gel상에서 전기영동을 통해 단편을 분리하여 분석한다. DNA의 변화가 이보다 작은 삽입에서는 크기 변화를 찾기 위해 DNA sequencing 또는 더 정밀한 전기영동 기술이 필요하다. 절단 위치에서의 역위와 재배열은 특정 영역 DNA를 절단하는 효소의 기능을 방해하여 해당 절단 위치에서 DNA를 자르지 못하게 하고 따라서 비교적 큰 DNA 단편의 변화를 일으키므로 검출하기 쉽다. 점 돌연변이는 단편 size의 변화를 일으키지 않기 때문에 검출하기가 더 어렵다.

수산생물 유전학에서는 몇몇 종류의 marker가

많이 이용되고 있다. 과거에는 isozyme과 mtDNA marker가 수산생물유전학 연구에서 많이 이용되었다. 현장에서 이용되고 있는 최근의 marker는 RFLP, RAPD, AFLP, microsatellite, SNP, 그리고 EST marker가 있다. 표 1은 이들 marker 종류의 기본적인 특성을 요약한 것을 나타냈으며, 각각의 marker에 대해서 자세히 설명하였다.

2.1 Type I vs Type II marker 그리고 다형성 정보 content (PIC)

분자 marker는 두 가지로 구분할 수 있다. Type I은 기능이 파악된 유전자와 관련된 marker이고, type II는 알려지지 않은 게놈 단편과 관련이 있다. 이런 구분에서는 대부분의 RFLP marker는

표 1. DNA marker의 종류, marker의 특성, 그리고 응용가능성

marker 종류	약어	사전에 필요한 분자정보	유전 방식 (Mode of inheritance)	종류	조사대상 유전자좌	추정되는 대립유전자의 수	다형성 또는 분해능력	주요 활용
Isozyme		○	멘델, 공우성	Type I	Single	2-6	낮음	Linkage mapping, population studies
Mitochondrial DNA	mtDNA	×	모계유전	-		다중	halotypes	Maternal lineage
Restriction fragment length polymorphism	RFLP	○	멘델, 공우성	Type I 또는 II	Single	2	낮음	Linkage mapping
Random amplified polymorphic DNA	RAPD AP-PCR	×	멘델, 우성	Type II	Multi	2	중간	Fingerprinting for population studies, hybrid identification
Amplified fragment length polymorphism	AFLP	×	멘델, 우성	Type II	Multi	2	높음	Linkage mapping, population studies
Micorsatellite	SSR	○	멘델, 공우성	거의 Type II	Single	다중	높음	Linkage mapping, population studies, paternity analysis
Expressed sequence tags	EST	○	멘델, 공우성	Type I	Single	2	낮음	Linkage mapping, physical mapping, comparative mapping
Single nucleotide polymorphism	SNP	○	멘델, 공우성	Type I 또는 II	Single	2개 그러나 4개까지	높음	Linkage mapping, population studies?
Insertions/deletions	Indels	○	멘델, 공우성	Type I 또는 II	Single	2	낮음	Linkage mapping

^a보존적 PCR primer는 근연종에서 염기서열정보로부터 채용될 수 있다(Conserved PCR primers can be adopted from sequence information from a related species)

type I에 해당하는데, 그것은 이들 marker는 밝혀진 유전자의 분석을 통해서 동정되기 때문이다. 또한 RAPD marker는 type II에 해당하는데, 그것은 RAPD band는 PCR을 통해 알려지지 않은 게놈 영역에서 증폭되었기 때문이다. AFLP marker 또한 미지의 게놈 영역에서 증폭되기 때문에 type II에 속한다. Microsatellite marker는 기능이 알려진 유전자와 연관이 없다면 type II에 속한다. EST marker는 유전자의 전사를 의미하므로 type I에 해당한다. SNP marker는 발견된 배열(eSNP 또는 cSNP)로부터 개발되지 않았다면 대부분 type II marker이다. Indel은 genomic 또는 transcriptomic sequencing project 과정에서 종종 발견된 후 marker로서 점점 더 널리 이용되고 있다. Indel (Insert와 deletion의 합성어)은 유전자내에 위치에 따라서 type I 또는 type II가 될 수 있다.

Type I marker의 의미는 비록 이들 marker가 매우 중요하다는 것은 확실해지고 있지만 수산생물 유전학의 초기 단계에서 충분히 알려지지 않았다. 집단 연구에서 marker로서 기능에 더하여 type I marker는 점점 유전적 연관 및 QTL 지도 작성을 위해 그 역할이 점점 증대되고 있다. Type I marker는 비교게놈학, 게놈진화, 후보 유전자 동정을 위해서 매우 유용하다. 게놈의 진화적 한계로 인해 많은 유전자와 이들의 특이성은 종내에 보존되어 있다. 비교게놈학은 게놈간의 유사성과 차이점을 검출하기 위해 유전적 정보가 매우 근연한 종에 대해 이용할 수 있다면, 수산생물 유전학 연구에 사용할 marker를 개발하는데 소요되는 많은 시간, 경비, 노력을 줄일 수 있다. 최근까지, 인간, 쥐, Zebra fish와 같이 많이 연구된 생물에서 얻은 정보에 의해 수산생물 게놈학에 대한 충분한 이해가 크게 좌우되고 있다. Type I

marker는 map-rich 종의 게놈 정보를 비교적 map-poor 종과의 비교 및 전달하기 위한 역할을 한다. 이런 종간 비교는 또한 type II marker를 통해서도 가능하지만, 비교 범위는 아주 근연한 분류군에 한정된다. 이런 비교를 위해 필요한 것은 배열의 보존에 있다. 가장 자주 이용되는 MS (microsatellite) marker에 있어서, 그런 비교 연구는 PCR primer 디자인을 위해 사용되는 반복배열의 양쪽에 있는 염기서열의 보존에 좌우된다. 대조적으로, 유전자내 배열의 보존율이 높으면 type I marker가 종간에 비교할 게놈 단편에 대한 견인차로서의 역할을 한다. 예를들어, 15개 유전자가 Zebra fish내 type I marker A와 B사이에 있다면, 비록 유전자의 정확한 수, 순서, 방향이 꼭 동일한 것은 아니지만 15개 유전자는 대부분 메기의 marker A와 B상에 존재할 가능성이 있다. 현재, 수산생물 유전학에서 주요 종에 대해서는 large insert bacterial artificial chromosome (BAC) library를 이용하고 있다. 이들 종에는 차넬메기, 틸라피아, 대서양연어, 무지개송어 등이 있다. 가까운 시일에 linear arrays of large insert clone이 BAC contigs로서 알려진 clusters of overlapping clone으로 지시될 것이며, BAC contigs로 관심 영역내(중요한 QTL을 가지는 것으로 추정) 많은 분자 marker의 개발이 가능할 것이다. 또한, 형질관련 marker 주위의 DNA 배열에 대한 추가적 분석으로 QTL을 나타내는 DNA의 동정도 가능할 것이다.

일반적으로 RAPD, MS, AFLP와 같은 type II marker는 유전정보를 포함하고 있지 않으며, 따라서 자연선택에 중립적인 것으로 판단된다. 이런 marker는 집단내 및 집단간 유전적 다양성과 유전적 분화의 규명이 Hardy-Weinberg 평형 그리고 이용된 marker가 선택중립성이라는 가정하에 집단

유전학 연구에서 널리 이용되고 있다(Brown and Epifanio, 2003). 또한 type II marker는 종, 계통, 그리고 잡종 동정을 위한 수산생물 유전학 및 육종연구에서 유용한 것으로 증명되었고, 더 최근에는 QTL과 연관된 marker로서도 유용한 것으로 나타났다.

분자 marker의 유용성은 다형성정보 함유량(PIC, Botstein et al., 1980)을 근거로 측정할 수 있다. PIC는 한 집단내 다형성을 검출하기 위한 marker의 유용성을 가리킨다. PIC는 검출할 수 있는 대립유전자의 수와 그 대립유전자 빈도의 분포에 좌우되고, 모든 대립유전자 빈도의 평방합에서 1을 뺀 값이다. 예를 들어, 빈도 0.5의 두 대립유전자를 가진 MS marker의 PIC는 각각 $1 - [(0.5)^2 + (0.5)^2] = 0.5$ 이고, 반면 빈도 0.9와 0.1의 두 대립유전자의 MS marker에 대한 PIC는 0.18이다. 대립유전자의 빈도가 크면 클수록 PIC는 커진다. 대립유전자의 수가 한정되어 있는 경우, 대립유전자의 빈도가 같으면 PIC는 더 커진다. PIC 값의 비교를 통해, 연구자들은 수산생물 유전학에서 연구해야 할 대상중에서 고려할 다양한 marker의 분해 능력을 대략 알 수 있다.

2.2 Isozyme marker

Isozyme은 단일 유전자좌에 의해서 생산된 대립유전자성 단백질 변이이며 isozyme은 다형성을 나타내고 유전자의 단백질 산물을 나타내므로 type I이며 marker로서 관심의 대상이다. 1960년대 이후로, isozyme의 starch gel 전기영동은 어업 관련 유전적 연구에서 가장 일반적으로 이용된 방법(Ryman and Utter, 1987; Hillis et al., 1996)이며, 아직도 여전히 많이 사용되고 있다.

한 효소의 대립유전자성 폴리펩타이드 체인의

아미노산 차이는 근간이 되는 DNA 배열의 변화를 의미한다. 아미노산 변화에 따라서, 전기가 흐르는 starch gel상에서 만들어진 단백질 산물의 이동 속도는 달라진다(전기하전도와 입자크기의 차이에서 기인). 대립유전자의 존재여부와 상대적 빈도의 차이는 유전적 변이를 정량화하고 집단, 종, 그리고 더 높은 분류적 목적에서 유전적 단위간을 구별하기 위해 이용된다. Isozyme은 근친도 추적, 계군동정, 친자분석을 위해 양식에서 이용된다. 몇 가지 사례에서는 일부 isozyme marker와 양적형질간의 상관관계가 보고되기도 했다(Hallerman et al., 1986; McGoldrick and Hedgecock, 1997). Linkage map 작성을 위한 활용은 연어류(Pasdar et al., 1984; May and Johnson, 1993), poeciliids (Morizot et al., 1991) 연구에서 증명되었지만, 이용 가능한 isozyme 유전자좌의 수가 한정되어 대규모 계통 지도 작성에서는 사용되지 못했다.

Isozyme과 관련된 단점은 null 대립유전자로 인한 이형접합체 결핍, 그리고 필요한 시료(조직)의 양과 질이다. 거기다 약간의 DNA 배열차이는 단백질 수준에서 나타나지 않아서, 검출 가능한 변이의 수준이 줄어든다. 뉴클레오타이드 배열에서의 약간의 변화는 encoded 폴리펩타이드를 변화시키지 않고, 일부 폴리펩타이드 변화는 전기영동 겔에서 단백질의 치환을 변화시키지 않는다. 비록 수 백개의 유전자좌를 나타내는 75개 isozyme을 이용할 수 있지만, 보통 이용되는 유전자가 비교적 적고, 대부분의 유전자좌에 의해 나타내어지는 대립유전자의 수가 적어서(보통 2~3) 이들 marker의 PIC를 매우 과소 평가하는 경향이 있다. 수산생물 집단에 대한 많은 isozyme 연구에서 나타난 낮은 유전적 변이(e.g. Siddell et al., 1980; Mork et al., 1985; Crawford et al., 1989)는 더 큰 유

전적 해상력을 가진 marker에 대한 연구를 촉진시켰다. 공우성 type I marker로서의 강점, 이용의 편리함, 그리고 낮은 비용에도 불구하고 양식유전학에서 이용은 제한되었다.

2.3 미토콘드리아 DNA marker

척추동물에 대한 연구에서 염기서열 분화는 핵 DNA보다 미토콘드리아 DNA에서 더 빠르게 축적되는 것으로 나타났다(Brown, 1985). 이것은 mtDNA가 복제과정에서 수복기작이 없어 돌연변이율이 더 높기 때문이고(Wilson et al., 1985), 반수체 미토콘드리아 게놈의 모계 유전으로 인해 유효집단 크기가 더 작기 때문이다(Birky et al., 1989). 복제와 전사를 시작하는 약 1kb control region(D-loop)을 제외하고 거의 전체 mtDNA 분자는 전사된다. 일반적으로 D-loop와 같은 비유전자 영역 단편은 cytochrome b 유전자와 같은 코딩 배열에 비해 상대적으로 많은 변이를 보이고(Brown et al., 1993), 이것은 아마도 기능적 제한의 감소와 선발압력의 완화에서 기인된 것으로 보인다.

mtDNA marker의 분석은 뱀장어(Avise et al., 1986), bluefish(Graves et al., 1992), red drum(Gold et al., 1993), snappers(Chow et al., 1993) 그리고 shark(Heist and Gold, 1999)를 포함하는 다양한 어류에서 집단구조를 조사하기 위해 광범위하게 사용되었다. 미토콘드리아 marker는 수산생물 유전학자에게 아주 많이 이용되는데, 이것은 부분적으로 친어 집단의 동정에 이용되기 때문이다 (e.g. Benzie et al., 2002). 분자 분석 초기에 isozyme에 비해 상대적으로 mtDNA의 다형성이 높아서 집단 분화의 연구를 위한 수산생물 유전학에서 활용되었다. 기술적으로 mtDNA marker는 목적 분자가 핵 게놈 DNA가 아니라 mtDNA라는(Cronin et al.,

1993) 점을 제외하고는 RFLP marker에 해당된다. 비록 mtDNA 유전자좌가 유전자좌당 대립유전자의 수가 많다고 할지라도, mtDNA 분자에서 이용 가능한 marker가 한정되어 있어서 isozyme 보다는 PIC 값이 높지만 RAPD, microsatellite, AFLP 그리고 SNP와 같은 변이성이 큰 핵 DNA marker보다는 낮다.

비텐델의 유전법칙에 따라 mtDNA 분자는 유전적 조사에서 하나의 유전자좌로서 고려되어야 한다 (Avise, 1994). 또한 mtDNA는 모계 유전되기 때문에, mtDNA에서 얻은 계통수와 집단구조는 gender-biased migration (Birky et al., 1986) 또는 이입 (Chow and Kishino, 1995)으로 인하여 핵 게놈의 변화를 반영하지 않을 수도 있다. 또한, mtDNA marker는 돌연 변이, 평행 치환 그리고 이형접합체율이나 mutational hot spot와 같은 다른 DNA에 기반을 둔 marker에서 존재하는 문제들이 나타난다.

2.4 Restriction fragment length polymorphism (RFLP)

RFLP marker (Botstein et al., 1980)는 생명과학에서 또 하나의 출발계기가 된 게놈 혁명의 첫 신포탄으로서 받아들여진다. RFLP marker의 과학적 절차와 원리는 Fig. 1에 요약되어 있다. 제한효소는 박테리아 효소로서, 이 효소는 특정 4,5,6 또는 8 염기를 인식해서 이들 배열이 나타날 때마다 DNA를 절단한다. 그래서 indel, 염기 치환 또는 제한지점과 관련된 DNA 배열의 변화로 절단 지점(Fig. 1)이 새로 생성되거나 없어지며 또는 위치가 바뀔 수 있다. 제한효소로 DNA를 절단시키면 DNA 단편이 생기는데, 이 단편의 수는 개체, 집단, 그리고 종에 따라 다양하게 나타난다. 전통적으로 DNA 단편은 Southern blot analysis를

이용해 분리하고(Southern, 1976), 이 분석에서 게놈 DNA를 분해하여 agarose gel에 전기를 걸어 단편을 막으로 이동시키고 특정 probe를 붙여서 시각적으로 확인한다.

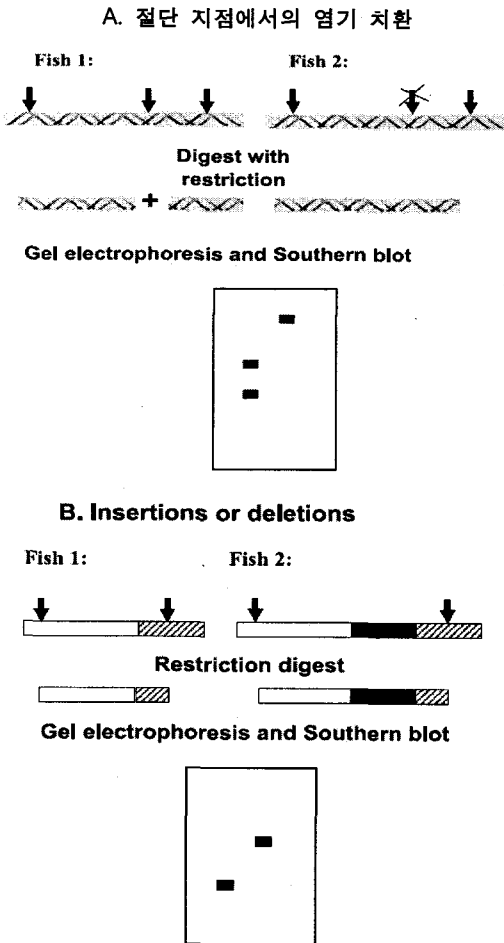


Fig. 1. Molecular basis of restriction fragment length polymorphism (RFLP). Loss of a restriction within the locus of interest leads to loss of one fragment and increase of fragment size that can be resolved by gel electrophoresis. (B) Insertion of a piece (black bar) between two restriction sites within the locus leads to an increase in the fragment size that can be resolved by gel electrophoresis (reversely a deletion should reduce the fragment size).

최근에는 PCR을 기반으로 한 기술을 이용해서 복잡한 Southern blot법을 대신한다. 한 유전자좌에 대한 양측의 염기서열을 알고 있다면, RFLP 영역을 포함하는 단편을 PCR을 통해 증폭한다. 길이의 다형성은 비교적 큰 제거나 삽입의 경우는 전기영동을 통해 그 차이를 구별한다. 그러나 길이 다형성이 제한 지점의 하나의 염기치환으로 유발된 것이라면, PCR 산물을 제한효소로 분해시켜야 RFLP를 확인할 수 있다. 최근에는 이용할 수 있는 universal primer의 수가 많아지고 있어서, 연구자들은 비교적 잘 보존되거나 빠르게 진화하고 있는 DNA 영역을 타겟으로 잡을 수 있고, 이것은 대상종에서 관찰된 변이의 양과 분류학적 위치에 따라 좌우된다. 더우기, PCR 산물은 제한효소로 분해할 수 있고, PCR을 통해 생산된 다량의 DNA 양은 EtBr로 염색해서 쉽게 확인할 수 있다.

유전적 변이를 찾는데 있어서 RFLP marker의 잠재력은 다음에서 언급될 더 최근에 개발된 marker나 기술에 비해 상대적으로 낮다. 제한지점을 포함하는 영역의 Indel과 재배열은 대부분의 생물의 게놈에서 넓게 퍼져있겠지만, 연구하고 있는 한 유전자좌내에서 Indel이나 재배열이 있을 가능성은 드물 것이다. 유사하게, 10^9 염기쌍으로 된 한 게놈에서, 6bp을 인식하는 한 제한효소에 대해 약 250,000 절단 지점이 존재한다(이것은 1.5×10^6 또는 전체게놈의 0.15%에 해당). 절단 지점내에서 염기치환은 광범위하게 존재할 것이지만, 또한 연구하고 있는 유전자좌에서 염기치환이 일어날 가능성은 비교적 적을 것이다.

RFLP marker의 주된 강점은 이들 marker가 공유성이라는 점이다. 즉, 한 개체내에서 두 개의 대립유전자가 나타난다. 가끔 크기의 차이가 뚜렷하지만 구별하는 것은 쉽지 않다. RFLP의 주요 단

점은 비교적 다형성이 낮다는 점이다. 더군다나 염기서열 정보 또는 Southern blot 분석을 위한 probe가 필요해서 분자정보가 부족한 종에서 marker를 개발하는 것이 어렵고 시간도 많이 든다.

2.5 Random amplified polymorphic DNA(RAPD)

RAPD 과정은 길이 8-10bp의 동정된 primer로 밝혀지지 않은 핵 DNA 영역을 무작위로 증폭시키기 위해 PCR이 이용되면서 1990년에 처음 개발되었다. primer는 짧고 비교적 낮은 annealing 온도(종종 36-40°C)가 이용되었기 때문에, 다중 산물이 증폭될 가능성은 많았고 이들 다른 산물로 인해 다른 유전자좌가 나타나게 된다. 척추동물의 핵 게놈의 대부분은 유전정보를 가지고 있지 않고, 증폭된 유전자좌의 대부분은 선발에 중립적이다. 대상 분류군내 그리고 분류군 사이의 유전적 변이와 분화는 primer에 의해 생산된 산물의 존재 또는 부재에 의해 평가되었고, 이들 산물은 각 유전자좌에서 DNA 배열의 변화에 의해 판정되었다. RAPD 다형성은 primer 결합 영역에서의 염기 치환 또는 이들 지점 사이의 영역의 indel에 의해 일어날 수 있다(Fig. 2). 다형성 조사에 있어서 잠재력은 비교적 크다. 보통 5-20 band가 하나의 primer 쌍을 이용해서 얻을 수 있고 다중 세트의 random primer는 여러 RAPD band에 대한 전체 게놈을 조사하기 위해 이용할 수 있다. 각 band는 bi-allelic locus로 여겨지기 때문에 (증폭된 산물의 존재 또는 부재), RAPD에 대한 PIC 값은 micro-satellite와 SNP의 PIC 보다는 낮고, RAPD는 동시에 생산되는 유전자좌가 적어서 AFLP만큼 많은 정보를 제공하지 못할 수도 있다.

RAPD marker는 우성의 멘델 marker로서 유전

되고 존재/부재로서 기록된다. 하나의 RAPD band는 이형접합체뿐만 아니라 동형접합체에 의해 생산되고, band 강도는 다를 수 있지만, PCR 효율성의 변이로 인해 band 강도의 구분이 어려워진다. 결과적으로 이형접합체 개체와 동형접합체의 우성 표현형을 구별하는 것은 일반적으로 불가능하다. 게다가, band가 다른 유전자좌 또는 한 유전자좌의 다른 대립유전자를 나타내는지 파악하기 어려워, 연구 대상의 유전자좌 수는 잘못 평가될 수 있다. 특히 RAPD가 primer 결합 영역보다는 유전자좌내 제거나 삽입에 의해 유발된다면 더욱 그렇다. 육종연구에서, F1 세대에서 나타난 RAPD band의 수는 부모에서 나타난 band의 합계와 같아야 한다. 이것은 각 유전자좌에서 부모와 동형접합체란 것을 가정한다. 이 경우에 다형성 RAPD는 F2 집단에서 3:1 비율로 분리된다(Fig. 3, Liu et al., 1998a, 1999b).

RAPD는 PCR 기반 marker의 모든 장점을 가지고 있을 뿐 아니라, primer가 상업적으로 이용 가

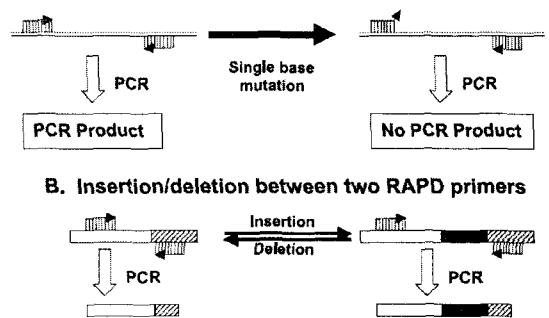


Fig. 2. Molecular basis of RAPD polymorphism. (A) Base substitutions in the primer binding sites, especially at the 3' end of the primer binding sites may lead to decrease (as shown) or increase of the number of RAPD bands. (B) Insertion or deletion between two primers may lead to increase or decrease of fragment sizes.

능하고 또한 대상 DNA 배열이나 유전자 구성에 대한 사전 지식이 없어도 되는 장점이 있다. 비록 더 높은 분해능은 비용이 더 들고 손이 많이 가는 방법이지만 불명확한 polyacrylamide gel 전기영동 (dPAGE)과 silver staining (Dinesh et al., 1995)으로 가능했지만, 다중 유전자좌 증폭은 전기영동을 통해 구분해서 EtBr 염색으로 구분가능하다. RAPD의 다른 장점은 많은 유전자좌 및 개체에 대한 검사의 편리함이다. RAPD marker는 어류(Partis and Wells, 1996) 그리고 연체류(Klinbunga et al., 2000; Crossland et al., 1993)에서 종 동정, black tiger shrimp (Tassanakajon et al., 1998)와 해조류(Van Oppen et al., 1996)의 집단구조 분석, 환경 스트레스의 유전적 영향분석(Bagley et al., 2001), 그리고 유전적 다양성 분석(Wolfus et al., 1997; Hirschfeld et al., 2002; Yue et al., 2002)에서 이용되었다.

이 유전적 marker의 단점은 유전자좌의 멘델유전 법칙을 증명하는데 어려움, 그리고 동형접합체

와 이형접합체 사이의 구분이 불가능하다는 점이 다. 끝으로, RAPD marker는 PCR 증폭과정에서 낮은 annealing 온도로 인한 재현성이 낮다.

2.6 AFLP (Amplified fragment length polymorphism)

AFLP는 앞서 논의한 RFLP와 RAPD 방법의 장점을 강화하고 단점을 극복하기 위한 PCR 기반 multilocus fingerprinting 기술이다(Fig. 4). RFLP와 유사하게, AFLP 다형성의 분자적 기초는 제한지점과 제한지점의 염기치환 사이의 indel을 포함한다. 마찬가지로 RAPD와 PCR primer 결합 영역에서의 염기 치환도 관련이 있다. 이 기술의 한 특징은 전체 게놈 DNA를 소화시켜서 생성한 DNA 단편에 이미 알려진 배열의 adaptor을 붙이는 것이다. 이렇게 해서 gel 전기영동에서 간편하게 분리하기 위해 전체 단편의 일부에 대한 PCR 증폭이 가능하다. 유전적 변이는 RFLP와 같지만, 한번에 하나의 유전자좌를 분석하는 것이 아니라 동시에 여러 개의 유전자좌를 분석할 수 있다.

이 방법은 Vos et al.(1995)에 의해 처음 사용되었고, AFLP 생산은 두 개의 효소(대부분 *Eco R I* 과 *Mse I*)로 전체 게놈 DNA를 소화시키면서 시작된다. 이렇게 만들어진 DNA 단편의 배열은 알 수 없으므로, 이 단편의 끝에 배열을 알고 있는 adaptor를 붙이고 PCR 증폭을 위한 primer 영역으로 이용한다. 이를 통해 수백만 개의 PCR 단편이 만들어지는데, 증폭된 단편의 수는 PCR primer의 3'에 염기를 붙여서 줄인다. 이들 염기는 접합 지점을 지나 DNA 단편으로 이어지므로, primer는 단편이 정확한 배열을 가질 경우에만 붙는다. 해당 DNA 부위는 4개의 염기 중 하나(A, T, G, 또는 C)를 가지므로, primer 하나에 하나의 염기를 붙

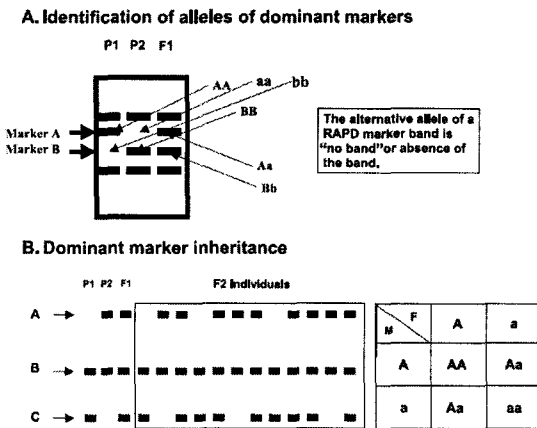


Fig. 3. Allele identification and inheritance of dominant markers. (A) Identification of alleles of dominant markers. For a mating between parent 1 (P1) and parent 2 (P2) to produce F1, two bands are different between the two parents and therefore can be followed as markers.

임으로 증폭된 단편의 양은 4배 줄어든다. 양쪽 primer의 하나의 염기를 붙임으로 단편은 16배 줄어들게 된다. 각 PCR primer에 3개의 염기를 붙이면 4,096배로 줄어든다. AFLP 분석에서, *Eco R I* primer에만 라벨이 붙는다. 이 효소로 DNA를 절단하면 10^9 bp의 게놈에서 약 250,000개의 단편 또는 대부분의 *Eco R I* 단편이 *Mse I*에 의해 추가적으로 절단되므로 총 500,000 *Eco R I*-*Mse I* 단편이 만들어진다. (*Mse I*은 4bp를 인식하고, 반면 *Eco R I*은 6bp를 인식하여 절단한다). 따라서, PCR primer에 3개의 염기를 추가함으로써 *Eco R I*-*Mse I* 단편은 평균 약 122개의 band로 줄어들어 (500,000/4096) gel 전기영동을 통해 판별하기에 적당한 수로 조절된다.

AFLP 분해능력은 게놈 다형성을 밝히는데 있어서 매우 뛰어나다. 삭제와 삽입, 그리고 primer 영역의 염기 치환을 포함하는 10^9 bp 게놈을 *Eco R I*-*Mse I*에 의해 절단해서 생성된 약 500,000 단편은 4,096 primer 조합의 AFLP scan으로 검사할 수 있다. 물론, 게놈을 조사하기 위해 다른 효소를 이용할 수 있으므로 분석 가능성은 무한하다. 가능한 경우, 모든 4,096 primer 조합을 통한 full scan은 매우 많은 시간이 소요되고 비용도 많이 들기 때문에 위에서 설명한 과정으로 만들어진 100개 정도의 다형성으로 대부분의 경우에 충분하다. 예를들어, Young et al. (2001)은 133개의 다형성 marker를 만들기 위해 AFLP 기술을 이용했고, 이중 23개는 무지개송어, cutthroat trout 및 이들의 잡종을 구분하는데 이용되었다(Young et al., 1996; Felip et al., 2000). 다른 활용 사례(high-resolution linkage maps의 작성 포함)는 Blears et al.(1998)의 총설에서 언급되었다.

RAPDs와 같이 AFLP marker는 우성 marker로서

유전된다. 비록 AFLP band의 공우성을 평가하기 위한 소프트웨어 프로그램이 존재하지만, 공우성 평가는 잘 파악된 가계를 이용할 때만 가능하고 개체군 연구에서는 어렵다. 이것은 AFLP gel상에서 대상의 개체 band를 따로 분리해 염기서열을 밝히고 각 band에 대해 유전자좌 특이적 primer를 만들어서 해결할 수 있다.

이것은 사실상 미지의 single-copy 핵 DNA marker를 생산하는 것이다. 비록 많은 시간이 소요되지만 전통적인 분석을 하면서 계산할 수 있는 하나의 이상적인 공우성 marker가 된다. 이런 marker는 캘리포니아 수역에서 고유종과 이식 송어 사이의 잡종을 조사하기 위해 이용되었다. 보존적 접근으로서 AFLP는 우성 marker로서 다루어져야 한다. 부모가 동형접합체라면 모계와 부계 부모에서 유래한 모든 band는 F1 세대에서 나타날 것이고, 반면 이형접합체 band는 분리될 것이

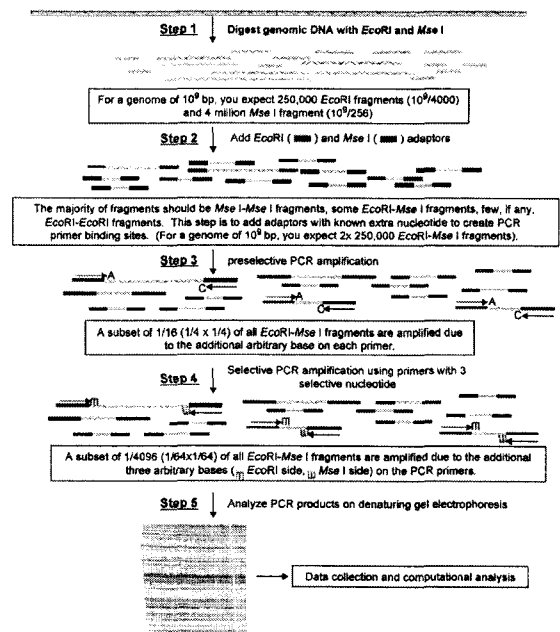


Fig. 4. AFLP 분석의 모식도.

다(Liu et al., 1998b, 1999c). AFLP 방법의 주요 강점은 많은 다형성 규명(100개 이상), 높은 PCR annealing temperature로 인한 높은 재현성, 그리고 marker당 상대적 경제성이다. RAPD 보다는 더 비용이 들지만 많은 수의 유전자좌를 동시에 분석할 수 있어서 marker당 비용은 크게 줄어든다. RAPD와 마찬가지로, 사전에 분자 정보가 필요하지 않고 따라서 어떤 종에도 적용할 수 있으며, 잘 연구되지 않은 종도 가능하다. 또한 RAPD 처럼 AFLP band는 비대칭 대립유전자인 것이 확인되어 상대적으로 낮은 PIC 값을 갖지만, 동시에 계산할 수 있는 유전자좌의 수가 많아서 유용성은 크게 증가한다. 주요 약점은 형광 label의 전기영동분석을 위해 자동화된 염기서열 분석기와 같은 특수한 장비가 필요하다는 것이다. 전통적인 전기영동법을 이용할 수 있지만 이 방법으로는 방사선 표지 또는 은 염색과 같은 특수한 염색기술이 필요하다.

2.7 Microsatellites

Microsatellite는 다수의 연결된 단순반복배열로 구성되어 있고 크기는 1에서 6개 염기로 되어 있다(예. ACG 또는 GATA; Tautz, 1989; Litt and Luty, 1989). MS는 최근까지 연구된 모든 종에서 풍부하고 어류에서 10kb 당 발견되는 것으로 추정되고 있다(Wright, 1993). MS는 모든 염색체와 염색체 내 모든 영역의 개체에 골고루 분포하는 경향이 있다. MS는 유전자 영역, 인트론, 그리고 비유전자 영역에서도 발견된다. 유전자 영역에서 가장 잘 알려져 있는 MS는 polyglutamine tract에 대한 정보를 가지고 있는 CAG 반복으로 이로 인해 정신박약이 일어난다. 대부분의 MS 유전자좌는 비교적 작고, 몇 개에서 수 백개에 이른다. MS

유전자좌의 크기가 상대적으로 작아서 PCR을 이용하는 genotyping에 중요하다. 일반적으로 말하자면 비록 다형성이 5반복 정도의 작은 반복을 가진 MS에서도 나타나지만, 많은 반복수를 가진 MS는 더욱 다형성을 띤다.

MS 다형성은 한 유전자좌에서 대립유전자가 가진 반복배열의 수의 차이에 의한 크기 차이에서 발생한다 (Fig. 5). MS 돌연변이율은 세대당 10^{-2} 만큼이나 높을 것으로 보고되었으며 (Weber and Wong, 1993; Crawford and Cuthbertson, 1996),

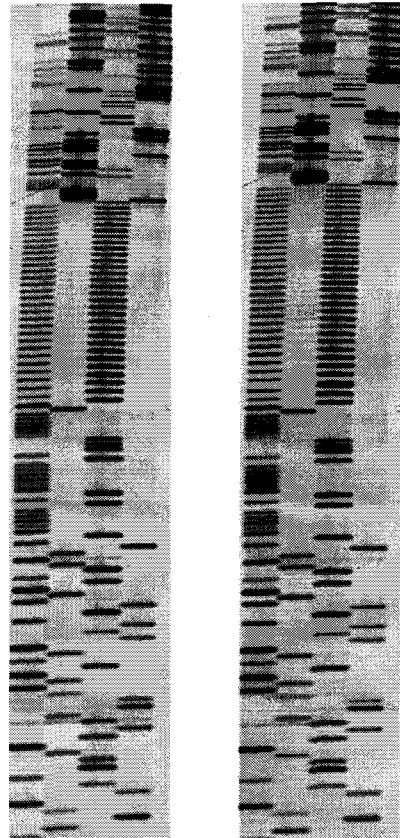


Fig. 5. The molecular basis of microsatellite polymorphism is the difference in repeat number among alleles numbers. Shown are sequence gels of two individual catfish that have identical flanking sequences, but differ by three AG repeats

DNA 복제시에 polymerase slippage에 의해 일어나는 것으로 여겨지는데, 이로 인해 반복배열의 수가 차이가 난다(Levinson and Gutman, 1987; Tautz, 1989). 인간 가계의 직접적인 연구에서는 새로운 MS 돌연변이는 보통 부모 대립유전자와 한 두번의 반복배열의 차이를 보이는 것으로 나타나(Weber and Wong, 1999), 순차적 돌연변이 모델을 지지한다(see review by Estoup and Cornuet, 1999).

그러나 몇몇 어종에서, 반복배열의 수에서 큰 차이를 보이는 대립유전자가 관찰되었고, 이것은 무한한 대립유전자 모델을 암시한다 (Balloux and Lugon-Moulin, 2002). 특정 기작과 상관없이, 반복배열의 수의 변화로 한 집단내에서 검출된 각 MS 유전자좌에서 대립유전자의 수가 증가될 수 있다.

MS는 공우성 marker처럼 멘델의 유전법칙에 따라서 유전된다(그림 6). 이것은 대립유전자의 풍부함, 전체 계놈상에 균등한 분포, 작은 유전자좌 크기, 그리고 높은 다형성과 더불어서 MS의 또 다른 강점이다. 그러나, MS marker를 이용하기 위해서는 많은 선행 투자와 노력이 요구된다. 각 MS 유전자좌가 파악되어야하고 PCR primer의 디

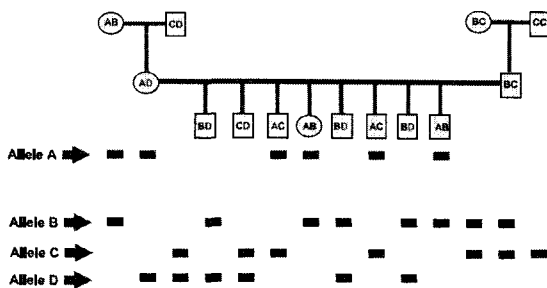


Fig. 6. Schematic presentation of co-dominant marker inheritance. Two pairs of matings were used to reduce the F1. In the first pair, the female has genotype AB at the locus; the male has genotype CD; and their F1 (one female individual) has genotype of AD.

자인을 위해 반복배열의 양측 영역의 배열도 알아야한다. 가장 효율적인 marker의 개발을 위해, MS가 풍부한 계놈 library가 만들어진다(Ostrand et al., 1992; Kijas et al., 1994).

복제과정에서 polymerase slippage로 인해, 해당 MS 유전자좌의 대립유전자 사이에 2bp 정도의 차이가 존재할 수 있다. 이 때문에, PCR 증폭된 MS DNA를 전통적으로 방사선으로 표지하여, sequencing gel에서 분리시키고 그 후 하룻밤 X-ray 필름 상에서 노출시킨다(Sambrook et al, 1989). 컴퓨터 이미지 시스템을 갖춘 자동화된 형광 sequencer를 이용해서, 하루에 판독할 수 있는 시료의 수도 크게 증가되었다(O'Reilly and Wright, 1995).

유전자좌당 대립유전자의 수가 많다는 것은 DNA marker의 PIC 값이 크다는 것이다. 광범위한 범위의 종에서 새롭게 규명된 MS 유전자좌에 대한 primer와 대립유전자 빈도를 발표하는데 노력하고 있는 잡지 *Molecular Ecology Notes* 의 최근 debut에서 증명된 것처럼, MS는 최근에 다양한 유전적 조사에서 아주 인기 있는 marker가 되었다. 지난 10년간에 걸쳐서, MS marker는 계놈지도 작성, 친자관계, 혈통관계, 집단구조를 포함한 어업연구에서 광범위하게 이용되었다(see O'Connell and Wright(1997) for review).

2.8 Single nucleotide polymorphism (SNP)

SNP는 한 유전자좌의 하나의 뉴클레오타이드 위치에서 다른 염기를 가지는 대립유전자를 생성하는 점 돌연변이에 의해 유발된 다형성을 설명한다. 염기의 치환에 의한 염기서열의 차이는 1977년 DNA sequencing을 시작한 후 잘 규명되었

지만, 많은 시료의 SNP를 빠르게 genotype하는 것은 1990년대 후반 gene chip 기술이 이용되고부터 가능했다. SNP는 어떤 생물내에 가장 풍부한 다형성을 띠고, 자동화가 가능하며, 다른 marker와 방법으로는 검출할 수 없는 숨겨진 다형성을 찾아낼 수 있어서 분자 marker 개발에서 다시 주목을 받고 있다.

이론적으로는 한 유전자좌내에 SNP는 4개의 대립유전자가 만들어질 수 있고 각각은 SNP 영역에 4개중 하나의 염기 즉 A, T, C, G의 하나를 가질 수 있다. 그러나 실제적으로 대부분의 SNP는 보통 두 개의 대립유전자 중 하나이고(종종 두 개의 pyrimidines C/T 또는 두 개의 퓨린 A/G) 비 대립유전자로 간주된다. 분명히 이들 PIC는 다중 대립유전자 microsatellite만큼 높지는 않지만, 이 단점은 변이의 풍부함으로 극복될 수 있다. SNP marker는 공우성 marker로서 유전된다.

SSCP 분석(Hecker et al., 1999), heteroduplex 분석(Sorrentino et al., 1992), 그리고 direct DNA sequencing을 포함한 여러 방법이 SNP discovery를 위해 이용되었으며, DNA sequencing은 가장 정확하고 많이 이용된 방법이었다. Random shotgun sequencing, PCR을 이용한 amplicon sequencing, comparative EST 분석은 SNP discovery를 위해 가장 인기 있는 sequencing 방법이다.

기술의 발전에도 불구하고 SNP genotyping은 여전히 어려운 일이고 특수한 장비가 필요하다. SNP genotyping을 위해 이용 가능한 전통적인 방법으로는 direct sequencing, single base sequencing(reviewed by Cotton, 1993), allele-specific oligonucleotide (ASO, Malmgren et al., 1996), deaturing gradient gel electrophoresis (DGGE, Cariello et al., 1988), single strand conformational polymorphism

assays(SSCP, Suzuki et al., 1990) 그리고 ligation chain reaction(LCR, Kalin et al., 1992)등이 있다. 각 방법이 장점과 단점을 가지지만, 모든 방법이 SNP genotyping에 여전히 유용하고 특히 예산과 인력이 부족한 실험실에서 더욱 그렇다. 그러나, 대량의 SNP marker 분석은 고가의 첨단 장비의 활용에 따라 좌우된다.

첨단 장비를 이용한 효율적 genotyping에는 여러 방안이 가능하다. 특히 인기있는 방법으로는 Matrix-assisted laser desorption ionization-time of light (MALDI-TOF) mass spectrometry (Ross et al., 1998; Storm et al., 2003), pyrosequencing(Ahmadian et al., 2000; Alderborn et al., 2000; He et al., 2003a,b), Taqman allelic discrimination(Li et al., 2004), real-time(quantitative) PCR(Nurmi et al., 2001), 그리고 microarray나 gene chip (Hacia et al., 1999)의 이용이 있다. Mass spectrometry와 microarray technology는 장비에 많은 투자가 필요하다. Pyrosequencing과 quantitative PCR 장비는 일반적으로 US\$ 100,000 이하이고, 수산생물 유전학 분야를 연구하는 많은 연구실에서 갖출 수 있을 것이다. 다른 참고 사항으로는 시료의 양과 관련한 genotyping의 비용이다. Microarray(gene chip) 기술과 정량적 PCR은 많은 수의 시료(유전자좌 당 수 천명)이 이용되는 의학과 임상 세트에서 특히 유용하고 이 때문에 gene chip과 hybridization probe의 개발에 소용되는 비용을 감당할 수 있다. Mass spectroscopy와 pyrosequencing은 양식연구에서 대부분의 경우처럼 비교적 작은 시료의 수를 취급할 때 (예. 유전자좌 당 수백 개체) 비교적 비용면에서 효과적이다. SNP genotyping에 관한 더 세부적인 정보에 대해서는 Vignal et al.(2002)의 최근 총설에 나와 있다.

2.9 Expressed sequence tags (ESTs)

ESTs는 cDNA 클론의 random sequencing으로부터 생성된 single-pass sequence 이다. EST 방법은 유전자를 동정하고 발현 양상을 통해 발현을 분석하는 가장 효율적인 방법이다 (Franco et al., 1995; Azam et al., 1996; Lee et al., 2000). 이 방법은 특정 생리학적 조건에서, 또는 특정 발달 단계에서 어떤 조직에서 발현된 유전자에 대해 빠르고 가치있는 초기 정보를 제공한다. EST는 발현된 유전자의 체계적 분석을 가능하게 하는 cDNA micorarray의 개발에 유용하다. 또한 게놈 mapping에서 이용된다(Boguski and Schuler, 1995; Hudson et al., 1995; Schuler et al., 1996).

Genome mapping에 있어서, EST는 소나 돼지와 같은 동물 게놈학에서 linkage map과 물리적 지도 작성을 위해 유용하고, radiation hybrid panel은 다형성이 없는 DNA marker의 지도 작성에 유용하다(Cox et al., 1990). Radiation panel은 hybrid cell의 line들로 구성되어 있고, 각 hybrid cell은 대상 종의 조사(照査) 염색체의 작은 단편을 가진다. 일반적으로, 대상 종에서 얻은 세포에 방사선을 조사해서 염색체를 작은 단편으로 분해한다. 방사선이 조사된 세포는 스스로 생존할 수 없다. 그러

나 이들 세포는 수용 세포와 융합해서 방사선이 조사된 작은 염색체 단편을 가지는 융합세포를 생성한다. 많은 융합세포내 염색체 분해 지점의 특징은 marker와 유전자의 연관 및 물리적 지도 작성이 가능해 진다. 포유류 게놈 지도 작성의 인기에도 불구하고 (Yang and Womack, 1998; Amaral et al., 2002; Korwin-Kossakowska et al., 2002; McCoard et al., 2002) radiation hybrid panel은 아직 수산생물에서는 이용할 수 없다. BAC library를 가지는 물리적 지도 작성이 훨씬 더 높은 분해능력을 제공하고 BAC library는 현재 많은 수산생물에서 이용할 수 있다는 점을 고려하면, 수산생물에서 radiation hybrid panel의 개발은 가까운 시일에는 어려울 것이다. 그러므로 EST는 다형성 EST가 동정되기만 한다면 양식종의 지도 작성에도 유용하다(Liu et al., 1999a). 또한 EST는 hybridization에 의해 물리적 지도에 나타낼 수 있고, 물리적 유전적 linkage map의 통합을 결국 EST가 linkage map에 연결시킬 수 있을 것이다. EST가 microsatellite와 연관된 것으로 나타나면 유전적 linkage map에 EST도 표시될 수 있다. 이런 면에서, microsatellite를 포함하는 EST는 type I marker의 풍부한 자원이다.

* 다음호에 2편이 게재됩니다(참고문헌 포함).